

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2001, 10:51:24 ; Search time 43.17 Seconds  
(without alignments)  
2370.223 Million cell updates/sec

Title: US-08-911-824-61  
Perfect score: 4622  
Sequence: 1 MIVTMRAMKRNKILYI.....GRGILNIPRRVQGFERSLL 873

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

- SPTREMBL\_15:\*
- 1: sp.archaea:\*
  - 2: sp.bacteria:\*
  - 3: sp.fungi:\*
  - 4: sp.human:\*
  - 5: sp.invertebrate:\*
  - 6: sp.mammal:\*
  - 7: sp.mhc:\*
  - 8: sp.organelle:\*
  - 9: sp.phage:\*
  - 10: sp.plant:\*
  - 11: sp.rodent:\*
  - 12: sp.virus:\*
  - 13: sp.vertibrate:\*
  - 14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3859	83.5	879	12 Q9WU9	Q9WU9 human immun
2	3670.5	79.4	876	12 Q9WU9	Q9WU9 human immun
3	3664.5	79.3	872	12 Q9WU9	Q9WU9 human immun
4	3664	79.3	871	12 Q9WU9	Q9WU9 human immun
5	3654.5	79.1	900	12 Q9WU9	Q9WU9 human immun
6	3650	79.0	871	12 Q9WU9	Q9WU9 human immun
7	3632	78.6	863	12 Q9WU9	Q9WU9 human immun
8	3577.5	77.4	860	12 Q9WU9	Q9WU9 human immun
9	3414	73.9	877	12 Q9WU9	Q9WU9 human immun
10	3381.5	73.2	876	12 Q9WU9	Q9WU9 human immun
11	2383	51.6	840	12 Q9WU9	Q9WU9 human immun
12	2370	51.3	863	12 Q9WU9	Q9WU9 human immun
13	2356.5	51.0	860	12 Q9WU9	Q9WU9 human immun
14	2356	51.0	862	12 Q9WU9	Q9WU9 human immun
15	2342.5	50.7	859	12 Q9WU9	Q9WU9 human immun
16	2338.5	50.6	537	12 Q9WU9	Q9WU9 human immun
17	2334.5	50.5	861	12 Q9WU9	Q9WU9 human immun
18	2334	50.5	859	12 Q9WU9	Q9WU9 human immun
19	2332.5	50.5	864	12 Q9WU9	Q9WU9 human immun

20	2330	50.4	870	12 Q9WU9	Q9WU9 human immun
21	2328.5	50.4	536	12 Q9WU9	Q9WU9 human immun
22	2328.5	50.4	845	12 Q9WU9	Q9WU9 human immun
23	2324.5	50.3	855	12 Q9WU9	Q9WU9 human immun
24	2324.5	50.3	861	12 Q9WU9	Q9WU9 human immun
25	2323.5	50.3	855	12 Q9WU9	Q9WU9 human immun
26	2321.5	50.2	851	12 Q9WU9	Q9WU9 human immun
27	2319.5	50.2	853	12 Q9WU9	Q9WU9 human immun
28	2317.5	50.1	856	12 Q9WU9	Q9WU9 human immun
29	2317	50.1	847	12 Q9WU9	Q9WU9 human immun
30	2316	50.1	854	12 Q9WU9	Q9WU9 human immun
31	2316	50.1	857	12 Q9WU9	Q9WU9 human immun
32	2315.5	50.1	868	12 Q9WU9	Q9WU9 human immun
33	2314.5	50.1	853	12 Q9WU9	Q9WU9 human immun
34	2314.5	50.1	857	12 Q9WU9	Q9WU9 human immun
35	2314.5	50.1	858	12 Q9WU9	Q9WU9 human immun
36	2314	50.1	854	12 Q9WU9	Q9WU9 human immun
37	2312.5	50.0	851	12 Q9WU9	Q9WU9 human immun
38	2312.5	50.0	854	12 Q9WU9	Q9WU9 human immun
39	2312	50.0	861	12 Q9WU9	Q9WU9 human immun
40	2310	50.0	861	12 Q9WU9	Q9WU9 human immun
41	2308	49.9	840	12 Q9WU9	Q9WU9 human immun
42	2308	49.9	846	12 Q9WU9	Q9WU9 human immun
43	2306.5	49.9	849	12 Q9WU9	Q9WU9 human immun
44	2305	49.9	854	12 Q9WU9	Q9WU9 human immun
45	2304.5	49.9	854	12 Q9WU9	Q9WU9 human immun

ALIGNMENTS

RESULT	1
Q9WU9	
ID	Q9WU9; PRELIMINARY; PRT; 879 AA.
AC	Q9WU9
DT	01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE	GP160 PRECURSOR...
GN	ENV.
OS	Human immunodeficiency virus type 1.
OC	Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX	NCBI_TaxID=11676;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-VI686;
RC	MEDLINE=92223950; PubMed=10207543;
RA	Janssens W., Heyndrickx L., Van der Auwera G., Nkengasong J.,
RA	Belnaert E., Vereecken K., Coppens S., Willems B., Franssen K.,
RA	Peeters M., Ndumbi P., Delaporte E., Van Der Groen G.,
RA	"Interpatient genetic variability of HIV-1 group O";
RT	AIDS 13:41-48(1999).
RL	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-VI686;
RC	MEDLINE=96426454; PubMed=8828748;
RA	Delaporte E., Janssens W., Peeters M., Buve A., Dibaba G.,
RA	Perret J.L., Ditsambou V., Georges Courbot M.C., Georges A.,
RA	Bourgeois A., Samb B., Henzel D., Heyndrickx L., Franssen K.,
RA	Van Der Groen G., Larouz B., Mbe J.R.,
RT	"Epidemiological and molecular characteristics of HIV infection in
RT	Gabon, 1986-1994";
RL	AIDS 10:903-910(1996).
RL	EMBL: X96526; CNA65373.1;
DR	INTERPRO: IPR000228;
DR	INTERPRO: IPR000777;
DR	PFAM: PF00516; GP120; 1.
DR	PFAM: PF00517; GP41; 1.
DR	Signal.
FT	SIGNAL
FT	CHAIN
FT	CHAIN
FT	POTENTIAL.
FT	GP120.
FT	132
FT	164

FT CHAIN 302 335 V3.  
 FT CHAIN 395 427 V4.  
 FT CHAIN 458 475 V5.  
 FT CHAIN 525 879 GP41.  
 SO SEQUENCE 879 AA: 99099 NM: BC6422ADDF1A3409 CRC64;

Query Match 83.5%; Score 3859; DB 12; Length 879;

Best Local Similarity 82.5%; Pred. No. 0;  
 Matches 730; Conservative 58; Mismatches 79; Indels 18; Gaps 8;

QY 1 MYTMRAMGRNRKRLGILYVALIIPCLSSQVATYAGVWEDAPVLEFCASDANTL 60  
 DB 1 MYTMRAMGRNRKRLGILYVALIIPCLSSQVATYAGVWEDAPVLEFCASDANTL 60  
 QY 61 TSTKHNWASQACVPTDTPPEHYLLNTNTDNFNTENTVEMOMODIISLWOSLPCV 120  
 DB 61 TSTKHNWASQACVPTDTPPEHYLLNTNTDNFNTENTVEMOMODIISLWOSLPCV 120  
 QY 121 QMTFICIONCCTDIKNNNTSGTENRST-----SEENMKCEPNTTVLADKREKCA 173  
 DB 121 QMTFICIONCCTDIKNNNTSGTENRST-----SEENMKCEPNTTVLADKREKCA 173  
 QY 121 QMTFICIONCCTDIKNNNTSGTENRST-----SEENMKCEPNTTVLADKREKCA 180  
 DB 121 QMTFICIONCCTDIKNNNTSGTENRST-----SEENMKCEPNTTVLADKREKCA 180  
 QY 174 LEYVSDLEFLADNNTNTNTLLINCSNTTIKQACPVSEPIPIYTCAPAGAIKCNCA 233  
 DB 174 LEYVSDLEFLADNNTNTNTLLINCSNTTIKQACPVSEPIPIYTCAPAGAIKCNCA 233  
 QY 181 LEYVSDLVK-TDN---STWYTLINCSNTTIKQACPVSEPIPIYTCAPAGAIKCNCA 236  
 DB 181 LEYVSDLVK-TDN---STWYTLINCSNTTIKQACPVSEPIPIYTCAPAGAIKCNCA 236  
 QY 234 FENGSGKSNISVYCTHGIKFTVSTQILNGLTSKIRIKNGNISDGNKIYVLSLSD 293  
 DB 234 FENGSGKSNISVYCTHGIKFTVSTQILNGLTSKIRIKNGNISDGNKIYVLSLSD 293  
 QY 237 FENGSGKSNISVYCTHGIKFTVSTQILNGLTSKIRIKNGNISDGNKIYVLSLSD 296  
 DB 237 FENGSGKSNISVYCTHGIKFTVSTQILNGLTSKIRIKNGNISDGNKIYVLSLSD 296  
 QY 294 IEITCVPRGNNOYEMKIGPMATSMALGTSNRSVAYCOYNTTWEKAKNAEAYL 353  
 DB 294 IEITCVPRGNNOYEMKIGPMATSMALGTSNRSVAYCOYNTTWEKAKNAEAYL 353  
 QY 297 ISMTCEPRGNH-TYOBKIGPMATSMALGTSNRSVAYCOYNTTWEKAKNAEAYL 355  
 DB 297 ISMTCEPRGNH-TYOBKIGPMATSMALGTSNRSVAYCOYNTTWEKAKNAEAYL 355  
 QY 354 ELINTEGNT-TMIPNRSODSDVEYTHLFHNGHEFFYCANTSEMFNTYLCNGTNCNT 412  
 DB 354 ELINTEGNT-TMIPNRSODSDVEYTHLFHNGHEFFYCANTSEMFNTYLCNGTNCNT 412  
 QY 356 ELVNTSNKTYVMIFNOSNDG-DPEVTHLFHNGHEFFYCANTSEMFNTYLCNGTNCNT 414  
 DB 356 ELVNTSNKTYVMIFNOSNDG-DPEVTHLFHNGHEFFYCANTSEMFNTYLCNGTNCNT 414  
 QY 413 QINSANG--MIPCKIKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNTTE 470  
 DB 413 QINSANG--MIPCKIKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNTTE 470  
 QY 415 QNNATHTDQIPCKIKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNTTE 474  
 DB 415 QNNATHTDQIPCKIKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNTTE 474  
 QY 471 NTFPRIGGDMKDIWENELFKYVAVKPSVAPPIAPVIGTGHRRKRAVGKMLFLG 530  
 DB 471 NTFPRIGGDMKDIWENELFKYVAVKPSVAPPIAPVIGTGHRRKRAVGKMLFLG 530  
 QY 475 TTFPRVGGMDIMTELEFKYVAVKPSVAPPIAPVIGTGHRRKRAVGKMLFLG 534  
 DB 475 TTFPRVGGMDIMTELEFKYVAVKPSVAPPIAPVIGTGHRRKRAVGKMLFLG 534  
 QY 531 VLSAAGSTMGAAATATVQVTHSVKIGVQOQDNILRAIOAOQELRLSYWGIROLBARL 590  
 DB 531 VLSAAGSTMGAAATATVQVTHSVKIGVQOQDNILRAIOAOQELRLSYWGIROLBARL 590  
 QY 535 VLSAAGSTMGAAATATVQVTHSVKIGVQOQDNILRAIOAOQELRLSYWGIROLBARL 594  
 DB 535 VLSAAGSTMGAAATATVQVTHSVKIGVQOQDNILRAIOAOQELRLSYWGIROLBARL 594  
 QY 591 ALETLIOQOOLNMGCGRLICTSVKMETWNTNTIN--QIWGNLTWQEQOQIDNV 648  
 DB 591 ALETLIOQOOLNMGCGRLICTSVKMETWNTNTIN--QIWGNLTWQEQOQIDNV 648  
 QY 595 ALETLIOQOOLNMGCGRLICTSVKMETWNTNTIN--QIWGNLTWQEQOQIDNV 654  
 DB 595 ALETLIOQOOLNMGCGRLICTSVKMETWNTNTIN--QIWGNLTWQEQOQIDNV 654  
 QY 649 SSTIYEEOKAOVOEOEKEKLEDEWASLMMWLDITKMTKIKIILYVALIGRIYV 708  
 DB 649 SSTIYEEOKAOVOEOEKEKLEDEWASLMMWLDITKMTKIKIILYVALIGRIYV 708  
 QY 655 SSTIYEEOKAOVOEOEKEKLEDEWASLMMWLDITKMTKIKIILYVALIGRIYV 714  
 DB 655 SSTIYEEOKAOVOEOEKEKLEDEWASLMMWLDITKMTKIKIILYVALIGRIYV 714  
 QY 709 MIVLNVNIRGQVOPSLQIPTRQOSEAETPGRTGEGGDEGRPLIPPOGLPLIY 768  
 DB 709 MIVLNVNIRGQVOPSLQIPTRQOSEAETPGRTGEGGDEGRPLIPPOGLPLIY 768  
 QY 715 MIVLNVNIRGQVOPSLQIPTRQOSEAETPGRTGEGGDEGRPLIPPOGLPLIY 774  
 DB 715 MIVLNVNIRGQVOPSLQIPTRQOSEAETPGRTGEGGDEGRPLIPPOGLPLIY 774  
 QY 769 DLRTIILMSYHLSTLIGTQVISHLGLMIGOKIIDACRICAAVHYWLOELOKSA 828  
 DB 769 DLRTIILMSYHLSTLIGTQVISHLGLMIGOKIIDACRICAAVHYWLOELOKSA 828  
 QY 775 DLRTIILMSYHLSTLIGTQVISHLGLMIGOKIIDACRICAAVHYWLOELOKSA 834  
 DB 775 DLRTIILMSYHLSTLIGTQVISHLGLMIGOKIIDACRICAAVHYWLOELOKSA 834  
 QY 829 TSLIDTFAVAVANTDDIIGIOLRGRIILIPRRVROGPRSL 873  
 DB 829 TSLIDTFAVAVANTDDIIGIOLRGRIILIPRRVROGPRSL 873  
 QY 835 TSLIDTFAVAVANTDDIIGIOLRGRIILIPRRVROGPRSL 879  
 DB 835 TSLIDTFAVAVANTDDIIGIOLRGRIILIPRRVROGPRSL 879

RESULT 2

QYKRB2 PRELIMINARY; PRT: 876 AA.  
 ID QYKRB2  
 AC QYKRB2  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE GP160 PROTEINOR.  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroviruses; Retroviridae; Lentiviruses.  
 OX NCBI\_Taxid=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CAV;  
 RX MEDLINE=99223950; PubMed=10207543;  
 RA Janssens W., Heyndrickx L., Van der Auwerda G., Nkengasong J.,  
 RA Belmans E., Verheijen K., Copens S., Williams B., Franssen K.,  
 RA Peeters M., Ndumba P., Delaporte E., Van der Groen G.,  
 RT "Interpatient genetic variability of HIV-1 group O";  
 RL AIDS 13:41-48 (1999).  
 DR EMBL: X96522; CA65306.1;  
 DR INTERPRO: IPR000328;  
 DR INTERPRO: IPR000777;  
 DR PRAM: PR00516; GP120; 1.  
 DR PRAM: PR00517; GP41; 1.  
 KW Signal.  
 FT SIGNAL. 1 29  
 FT CHAIN 30 524  
 FT CHAIN 132 156  
 FT CHAIN 158 199  
 FT CHAIN 301 335  
 FT CHAIN 394 425  
 FT CHAIN 458 474  
 FT CHAIN 525 876  
 FT CHAIN 876 AA: 98879 NM: B064D1B2EC656C9F CRC64;  
 SO SEQUENCE

Query Match 79.4%; Score 3670.5; DB 12; Length 876;  
 Best Local Similarity 80.1%; Pred. No. 1.5e-295;  
 Matches 707; Conservative 63; Mismatches 96; Indels 17; Gaps 12;

QY 1 MYTMRAMGRNRKRLGILYVALIIPCLSSQVATYAGVWEDAPVLEFCASDANTL 60  
 DB 1 MYTMRAMGRNRKRLGILYVALIIPCLSSQVATYAGVWEDAPVLEFCASDANTL 60  
 QY 61 TSTKHNWASQACVPTDTPPEHYLLNTNTDNFNTENTVEMOMODIISLWOSLPCV 120  
 DB 61 TSTKHNWASQACVPTDTPPEHYLLNTNTDNFNTENTVEMOMODIISLWOSLPCV 120  
 QY 121 QMTFICIONC-TDINNNTSGTENRSTSEENMKCEPNTTVLADKREKQALFYVD 179  
 DB 121 QMTFICIONC-TDINNNTSGTENRSTSEENMKCEPNTTVLADKREKQALFYVD 180  
 QY 180 LFTLANNTNT--TWYTLINCSNTTIKQACPVSEPIPIYTCAPAGAIKCNCAEEN 236  
 DB 180 LFTLANNTNT--TWYTLINCSNTTIKQACPVSEPIPIYTCAPAGAIKCNCAEEN 236  
 QY 181 LFTLANNTNT--TWYTLINCSNTTIKQACPVSEPIPIYTCAPAGAIKCNCAEEN 238  
 DB 181 LFTLANNTNT--TWYTLINCSNTTIKQACPVSEPIPIYTCAPAGAIKCNCAEEN 238  
 QY 237 GTGKSNISVYCTHGIKFTVSTQILNGLTSKIRIKNGNISDGNKIYVLSLSDIET 296  
 DB 237 GTGKSNISVYCTHGIKFTVSTQILNGLTSKIRIKNGNISDGNKIYVLSLSDIET 296  
 QY 239 GTGKSNISVYCTHGIKFTVSTQILNGLTSKIRIKNGNISDGNKIYVLSLSDIET 357  
 DB 239 GTGKSNISVYCTHGIKFTVSTQILNGLTSKIRIKNGNISDGNKIYVLSLSDIET 357  
 QY 297 ACVRPGRNNOYEMKIGPMATSMALGTS-NRSVAYCOYNTTWEKAKNAEAYL 355  
 DB 297 ACVRPGRNNOYEMKIGPMATSMALGTS-NRSVAYCOYNTTWEKAKNAEAYL 355  
 QY 356 INTBENTMTNRSODSDVEYTHLFHNGHEFFYCANTSEMFNTYLCNGTNCNTOSI 415  
 DB 356 INTBENTMTNRSODSDVEYTHLFHNGHEFFYCANTSEMFNTYLCNGTNCNTOSI 415  
 QY 416 NSANG--MIPCKIKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNT-ENT 472  
 DB 416 NSANG--MIPCKIKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNT-ENT 472  
 QY 416 TNDNGTMDIPCKIKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNT-ENT 475  
 DB 416 TNDNGTMDIPCKIKOVYNSWMMGSGGLYAPPIPGNLTCISHTGMILOMDAPWNT-ENT 475

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.  
GENCORE version 4.3

OM protein - protein search, using sw model

Run on: January 25, 2001, 10:51:25 ; Search time 15.7 Seconds  
(without alignments)  
1795.716 Million cells used

Title: US-08-911-824-61

Perfect score:

Sequence: 1 MIVTMRAMGKRRNRKLGILYI.....GRGILNIPRRVRQGFERSLL 873

Scoring table: BLOSUM62

Scoring scale: 0.00002  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

FOSSC-processing: Minimum Match 0%  
Maximum Match 100%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %	Query			DB	ID	Description
		Score	Match	Length			
1	2318.5	50.2	854	1	ENV_STVCZ	P17281 chimpanzee	
2	2250	49.5	856	1	ENV_HV12H	P05881 human	
3	2287	49.5	853	1	ENV_HV1EL	P04581 human	
4	2264.5	49.0	859	1	ENV_HV1MA	P04583 human	
5	2263.5	49.0	855	1	ENV_HV1Z6	P04580 human	
6	2261	48.9	863	1	ENV_HV1Z8	P05882 human	
7	2259	48.9	847	1	ENV_HV1S1	P19550 human	
8	2247.5	48.6	868	1	ENV_HV1C4	P05879 human	
9	2247	48.6	843	1	ENV_HV1Y2	P35961 human	
10	2242	48.5	853	1	ENV_HV1Z2	P12487 human	
11	2239.5	48.5	852	1	ENV_HV1S3	P19549 human	
12	2238.5	48.4	848	1	ENV_HV1JR	P20871 human	
13	2233.5	48.3	861	1	ENV_HV1KB	P31819 human	
14	2231.5	48.3	856	1	ENV_HV1L1	P31872 human	
15	2228.5	48.2	852	1	ENV_HV1BN	P12488 human	
16	2214.5	47.9	847	1	ENV_HV1W2	P05880 human	
17	2212.5	47.9	846	1	ENV_HV1ND	P18799 human	
18	2211.5	47.8	856	1	ENV_HV1B1	P03375 human	
19	2211.5	47.8	856	1	ENV_HV1MI	P05877 human	
20	2209.5	47.8	856	1	ENV_HV1PY	P03376 human	
21	2209	47.8	851	1	ENV_HV1B8	P04582 human	
22	2208.5	47.8	856	1	ENV_HV1H2	P04578 human	
23	2207	47.7	867	1	ENV_HV1J3	P1489 human	
24	2205.5	47.7	856	1	ENV_HV1L3	Q70626 human	
25	2205.5	47.7	861	1	ENV_HV1OY	P03377 human	
26	2197	47.5	855	1	ENV_HV1OY	P02888 human	
27	2194.5	47.5	856	1	ENV_HV1S3	P05878 human	
28	2192.5	47.4	856	1	ENV_HV1H3	P04624 human	
29	2190	47.4	865	1	ENV_HV1RH	P04579 human	
30	2189	47.4	855	1	ENV_HV1A2	P03378 human	
31	2180.5	47.2	853	1	ENV_HV1MF	P19551 human	
32	1274	27.6	885	1	ENV_STV2A	P12492 simian	
33	1269.5	27.5	856	1	ENV_HV2N2	P05883 human	

pending ~~at~~ removal

## ALIGNMENTS

**RESULT** 1

[illegible]

FT	CARBOHYD	336	336	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	351	351	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	384	384	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	392	392	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	426	426	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	432	432	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	446	446	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	450	450	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	601	601	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	608	608	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	628	628	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	854 AA;	95803 MW;	2E249AFAD4F29B3 CRC64;	
Query Match					
Best Local Similarity 50.2%; Score 2318.5; DB 1; Length 854;					
Matches 452; Conservative 148; Mismatches 235; Indels 53; Gaps 16;					
QY	5	MRAMGRNR---	KLGILYIV-MALIIPCLSSQLSYATYAGVPVWEDAAAPVLFASDANL	60	
DB	1	MKYMEKKRDNWSLSITIIITILLPCL-TSELWTVYVGVVWEDADPVLFASDAKA	59		
QY	61	TSTKHNWASQACVPTDTPHEYLNTVDNFINWENTAVQMQEDIIISLWQSLKPCI	120		
DB	60	HSTEAHNINATQACVPTDPSQVFLPNVIESFNMMKNMVDQMHEDIISLWQSLKPCV	119		
QY	121	QMTFMCIQMCTDKNNNTSGTNRSSSENPMKTCFEFNITTVLKDKKQKQALFYVSDL	180		
DB	120	KLPLCVLTQCSKANFSQAKLNTQSSPPLEMKNSFNVTTELKDKKQVYSLFYVEDV	179		
QY	181	TKLADNNNTMTTLNCNSTTIKQAPKVSFPIPIYCAPAGYALFKNSAEFNGTK	240		
DB	180	VNLGNENNT---YRLNCNTAITQACPKTSFPIPIHYCAPAGFALLKNDKDFSGKG	236		
QY	241	CSNISVVTCHGIPKVPSTQILINGLSKEKIRMGKNISDSKNIIVTLSSDIETCVR	300		
DB	237	CTNVSTVHC7HQAQPVVTTQLLINGSIAEGNITVRVENSKNTDVMVQLVEAVSLNCR	296		
QY	301	PGNNOVQEMKIGP-WAIVSMALGTSGNSRSRVAYCQVNTTEWELKAKNTAERYLELINT	359		
DB	297	PGNN-TRGEVOIGPMTFYNIENVGDTRS--AYCKINGTWTWRTVEEVYKALATSSNT	353		
QY	360	EGNTTIFNRSDQSDVEVTHLFHNGCHGEFFCYNTSEFNITFLCNGTNCNNTQNSAN	419		
DB	354	AAITL--NRA-SGGDPVETHHMFNGCGEFFCYNTSOIF-----TDNITNGI	397		
QY	420	GMIPCKLKQVRSWMRGGSLYAPPIGNLTCISHTGMLQMDAP-WNKTEN-TRPRIG	477		
DB	398	IILPCIRIQIVSSWVRGRIYAPPIRGNITCSNITGLLTSDTPVTNNSGNLTPRPTG	457		
QY	478	GMKMDLWRNELFYKYVRVKPFSVAPTPARPVIGTGREKRAV-GLGLMFLGVLSAAG	536		
DB	458	GNMKDLWRSELYKYVRIEPLSVAPTKARRHTVAKQDKRAAFGLGALFLGFAGAG	517		
QY	537	STMGAAATALTVQTHSVIKIGVQQQNLRAIQAOELLRLSVWGIROLRLALETLLI	596		
DB	518	STMGAAAVTLTVQARQLLSGIVQQNNLLKAEQAQHLQLSLWSYKQLQARLAVERYL	577		
QY	597	QNOQLNLWCKGRLCYTSVKWNETWRNTNINQINGNLTWQEDWQIDNVSSTIYEI	656		
DB	578	QDOQILGLWCGSKAVCYTTPVNNNSWPGSNSTDDIWNLTWQWDLVSNYTKIFGLL	637		
QY	657	OQAQVOEQNEKKLLEDEWASLNNWLDITKWLWYIKIALIIVGALIGVRIVMVLNVR	716		
DB	638	EEAQSOEKNERDLELDQWASLNNWFDITKWLWYIKIFLMAVGGLIGRILMTVFSVR	697		
QY	717	NIRQGYQLSLQIPTRQQAETPGRTGGGGEGRPRIIPSPQGLFLYTLRLTIILW	776		
DB	698	RVQGYSPLSQTLIPVQREQGLGEIDEGGEGDRSRVRLVEGCLPLWDDLRLGIW	757		
QY	777	SYHLLNSLSIGTQTVISHURLGLW----	ILGQKIIDACR-----	ICAAVHYWLQELQ	825

DB	758	SYOSLTS-----LACNVWRQLTKLGHLSLRRLRRLCLLGGIIQYWGKELK	806	
QY	826	KSATSLIDTFAVAVANWTDIILGIQRLGRGILINIPRRVROGPERILL	873	
DB	807	ISALSUDATAIAVAGTDRIIEAFQVTLIRINPRIRQGLERALL	854	
RESULT 2				
ENV_HV12H	STANDARD;	PRT;	856 AA.	
AC	P05881;			
DT	01-NOV-1988 (Rel. 09, Created)			
DT	01-NOV-1988 (Rel. 09, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	ENVELOPE POLYPROTEIN GP160 PRECURSOR (CONTAINS: EXTERIOR MEMBRANE			
DE	GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)).			
GN	ENV.			
OS	Human Immunodeficiency virus type 1 (Zaire H321 isolate) (HIV-1).			
OC	Viruses; Retroviral viruses; Retroviridae; Lentivirus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-89228766; PubMed-2713163;			
RA	Srinivasan A., York D., Butler D. Jr., Jannoun-Nasr R., Getchell J.,			
RA	McCormick J., Ou C.Y., Myers G., Smith T., Chen E.;			
RT	"Molecular characterization of HIV-1 isolated from a serum collected			
RT	in 1976; nucleotide sequence comparison to recent isolates and			
RT	generation of hybrid HIV.;"			
RL	AIDS Res. Hum. Retroviruses 5:121-129(1989).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M15896; AAB53948.1; .			
DR	PIR; A44963; A44963.			
DR	HIV; M15896; ENV32321.			
DR	INTERPRO; IPR000328; .			
DR	INTERPRO; IPR000777; .			
DR	PFAM; PF00516; GP120; 1.			
DR	PFAM; PF00517; GP41; 1.			
KW	AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;			
KW	Signal.			
FT	SIGNAL	1	29	
FT	CHAIN	30	511	EXTERIOR MEMBRANE GLYCOPROTEIN.
FT	CHAIN	512	856	TRANSMEMBRANE GLYCOPROTEIN.
FT	DISULFID	53	73	BY SIMILARITY.
FT	DISULFID	118	206	BY SIMILARITY.
FT	DISULFID	125	197	BY SIMILARITY.
FT	DISULFID	130	153	BY SIMILARITY.
FT	DISULFID	219	248	BY SIMILARITY.
FT	DISULFID	229	240	BY SIMILARITY.
FT	DISULFID	297	331	BY SIMILARITY.
FT	DISULFID	379	445	BY SIMILARITY.
FT	DISULFID	386	418	BY SIMILARITY.
FT	CARBOHYD	87	87	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	132	132	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	152	152	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	156	156	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	198	198	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	242	242	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	263	263	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	277	277	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	294	294	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	302	302	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	332	332	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	339	339	N-LINKED (GLCNAC. . .) (POTENTIAL).



[illegible]

FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 618 618 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 819 819 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 859 AA; 97109 MW; DBCF9AA52E3ABF29 CRC64;  
  
Query Match 49.0%; Score 2264.5; DB 1; Length 859;  
Best Local Similarity 50.9%; Pred. No. 8.9e-157;  
Matches 442; Conservative 159; Mismatches 225; Indels 43; Gaps 19;  
  
QY 20 IVMALIIPCLSSOLYATYAGVYEDADAPVFCASDANLTSTKHNWASQACVPTDP 79  
DB 19 MLLGLMTCSTIAEDLWTVYGVPMKEATTLFCASDAKSYETEVHINWATHACVPTDP 78  
QY 80 TPBYLLTNVTDNINWYVMOEDIIISLWQSLKPCIQMTFMCIONCTDIKN-- 136  
DB 79 NPQIELENTVEGFMNKNWQHEHDIISLWQSLKPCVCLPLVTLNCTVNGTAV 138  
QY 137 NNTSGTENRTSS-----SENPKTCFEFNTVTKDKKQALFYVSDILKLANNTNTM 192  
DB 139 NGTNGSNRTNAELKMEIGVKNCSEFNTVPGSDKQ- EYATFYNDLIVQDDSD- NSS 195  
QY 193 YTLNCSNTIKQACPKVSEPIPIYCAPAGYAFICNSAEFGNGKCSNISVVTCTHG 252  
DB 196 YRLNCSNTVITQCPKVTDPPIHYCAPAGFAILCKDKFKNGTEICKNVSTVQCTHG 255  
QY 253 IKPVSTQILLNGSLSEKIRKMGKNTSDGKNIIVLTSDDIETTCVPRGNNQVQEMKI 312  
DB 256 IKPVSTQILLNGSLAEETMIRENTDNTKNIIVQLNETVTCNTPRGNN-TRRGHF 314  
QY 313 GP-NMAYSMALGTGNSRVAICYNTTEWKALEKNTAERYLEINNTGNTMIFNRSQ 371  
DB 315 GPQALYTTGI-VGDIRR--AYCINETENDKTLQYAVKLGSLNKT-----IIFN-SS 366  
QY 372 DGSVETHLHFNCHGEFFTCNTSEMYTFLNGTN-CNNTQINSANGMIPCKLKQV 430  
DB 367 SGGDPETHHSFNGRGEFFTCNTSKFNSTWQNNGARLSNSTESTGSI--LPCKRIQII 424  
QY 431 RSWMRGGGLYAPPIGNLTCISHTGIMLOMDAPMKNTE---TPRPIGDMKDIWRN 486  
DB 425 NMWQKTGKAMYAPPIAGVINCLSNITGLILTRDG-GNSNSDNETLPRGGDMRDWNIS 483  
QY 487 ELFKYKVVVRKVPFVAPPTIARPIVGTGTHREKRAVGLGMLFLGVLGAAGSTMGAATA 546  
DB 484 ELYKYKVVRIEPLGVAPTKAKRRV---EREKRAIGLGAFLGFLGAGSTMGAASITL 539  
QY 547 TVQTHSVKIGVQOQDNLRAIQAOQELLRLSVGIRQLRALLETIONQOLLNMG 606  
DB 540 TVQARQLSGVQOQDNLRAIEAQHLLQTLTVGIRQLQARVLAVERYLQDRLGMMG 599  
QY 607 CKGRICVTSVKNWRTNTNINOINGNLTWQEMDQIDNVSTIYEIOKAQVQEQN 666  
DB 600 CSGRHICTFPVWSSWSN-RSLDDINWNTMWMQWEKISNYITGLINLEESIQOEKN 658  
QY 667 EKKLEDEWASLWNLDTKWLVIKAIIVGALIGVRIVMIVLVRNIRQGYOPLS 726  
DB 659 EKELLEDKWASLWNSISKWLWYRIFIVVGLGILGRIFAVLSVNRVQGYSPLS 718  
QY 727 LQ--IPTRQSEATPGRTGEGGDEGRPLIPSPQGLPLLYTDLRTIILWSYHLSNL 784  
DB 727 LQ--IPTRQSEATPGRTGEGGDEGRPLIPSPQGLPLLYTDLRTIILWSYHLSNL 784

DB 719 LQTLPT-PRGPPDRPGIEBEGEGQGRSIRLVNFSALIWDLNLCFLSVHRLDL 777  
QY 785 ISGTQTVISHRLGLWILGQIKIDACRICAACAAVHYWQELQKQKATSLIDTFAVAVANWTD 844  
DB 778 LLIAITRIVE-----LLGRRGWEALKYLWNLLOYWQELKNSAISLLNTTAIAVACTD 830  
QY 845 DIILGIGRLGRLGILNIPRRVQGFERSLL 873  
DB 831 RVIEIGQRFGRAILHPIRRIRQGFERALL 859  
  
RESULT 5  
ENV\_HV126  
ID ENV\_HV126 STANDARD; PRT; 855 AA.  
AC P04580;  
DT 13-AUG-1987 (Rel. '05, Created)  
DT 13-AUG-1987 (Rel. '05, Last sequence update)  
DT 15-JUL-1999 (Rel. '98, Last annotation update)  
DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE  
DE GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].  
GN ENV.  
OS Human immunodeficiency virus type 1 (Zaire 6 isolate) (HIV-1).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
RN [J]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87248037; PubMed=3036660;  
RA Srinivasan A., Anand R., York D., Ranganathan P., Feorino P.,  
RA Schochetman G., Curran J., Kalyanaram V.S., Luciw P.A.,  
RA Sanchez-Pescador R.;  
RT "Molecular characterization of human immunodeficiency virus from  
RT Zaire: nucleotide sequence analysis identifies conserved and variable  
RL domains in the envelope gene.";  
RL Gene 52:71-82(1987).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DB EMBL; K03458; AAA5380.1; -  
DR PIR; D26192; VCLJZR.  
DR HIV; K03458; ENV526.  
DR INTERPRO; IPR000328; -  
DR INTERPRO; IPR000777; -  
DR PFAM; PF00516; GP120; 1.  
DR PFAM; PF00517; GP41; 1.  
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;  
KW Signal.  
FT CHAIN 1 30  
FT CHAIN 31 510 EXTERIOR MEMBRANE GLYCOPROTEIN.  
FT CHAIN 511 855 TRANSMEMBRANE GLYCOPROTEIN.  
FT DISULFID 53 73 BY SIMILARITY.  
FT DISULFID 118 207 BY SIMILARITY.  
FT DISULFID 125 198 BY SIMILARITY.  
FT DISULFID 130 155 BY SIMILARITY.  
FT DISULFID 220 249 BY SIMILARITY.  
FT DISULFID 230 241 BY SIMILARITY.  
FT DISULFID 298 332 BY SIMILARITY.  
FT DISULFID 378 444 BY SIMILARITY.  
FT DISULFID 385 417 BY SIMILARITY.  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).



QY 654 SONEKKLELDEWASLWNMLDITKWLWYIKIAIIIVGALIGVRIWVLNVRNQGYQ 720  
Db 653 EKNEQELLEDKKASLWNWENITQWLWYIKIFIMVGGILGLRIVFVLSLWNRVQGY 712  
QY 724 PLSQIPTTROQSAEPTPGTGGGDEGRPHLIPSPQGFPLYTDLRTILWYSYHLN 783  
Db 713 PLSQTLLPAPREDPREGIEEGEGGERDRSIRLVNGFSALIWDDLNLCLFSYHRLD 772  
QY 784 LLSGTQTVSHLRLGLWILQRIIDACRICAAVTHYLQELQKSATSLDITFAVAVNT 843  
Db 773 LLTAAARIVE-----LLGREGWEALYNLWILQYSRELRNSASLLDTTIAVAEGT 825  
QY 844 DDILIGLQRLGRGILNIPRRVQGFERSLL 873  
Db 826 DRVIEIVRTYRAVLNVPTRIQGLERLL 855

RESULT 6  
ENV\_HV128  
ID ENV\_HV128 STANDARD; PRG; 863 AA.  
AC P05882;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE ENVELOPE POLYPEPTIDE GP160 PRECURSOR (CONTAINS: EXTERIOR MEMBRANE  
GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)).  
GN ENV.  
OS Human immunodeficiency virus type 1 (Z-84 isolate) (HIV-1).  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88281278; PubMed=3395517;  
RA Yourho J., Josephs S.F., Reitz M.S. Jr., Zagury D., Wong-Staal F.,  
RA Gallo R.C.;  
RT "Nucleotide sequence analysis of the env gene of a new Zairian  
RT isolate of HIV-1";  
RL AIDS Res. Hum. Retroviruses 4:165-173(1988).  
CC -!- MISCELLANEOUS: THE Z-84 ISOLATE WAS TAKEN FROM A 54 YEAR-OLD  
CC ZAIREAN MALE.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announc>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL: J03653; AAA44684.1; -  
DR HIV: J03653; ENV\$YJ1.  
DR INTERPRO: IPR000328; -  
DR INTERPRO: IPR000777; -  
DR FRAM: PF00516; GP120; 1.  
DR FRAM: PF00517; GP41; 1.  
KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;  
KW Signal.  
FT SIGNAL  
FT CHAIN 1 29  
FT CHAIN 30 518  
FT CHAIN 519 863  
FT DISULFID 53 73  
FT DISULFID 118 217  
FT DISULFID 125 208  
FT DISULFID 130 160  
FT DISULFID 230 259  
FT DISULFID 240 251  
FT DISULFID 308 342  
FT DISULFID 388 452  
FT DISULFID 395 425  
FT CARBOHYD 87 87  
FT CARBOHYD 129 129  
FT CARBOHYD 136 136  
FT CARBOHYD 142 142  
FT CARBOHYD 143 143

-----  
FT EXTERIOR MEMBRANE GLYCOPROTEIN.  
FT TRANSMEMBRANE GLYCOPROTEIN.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT BY SIMILARITY.  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).



Query Match 48.98; Score 2261; DB 1; Length 863;  
 Best Local Similarity 49.5%; Pred. No. 1.6e-156;  
 Matches 442; Conservative 152; Mismatches 245; Indels 54; Gaps 15;

QY 5 MRAMGKRRN-----KLGILYVWALIIPLSSQLYATYAGVPWEDAAPVLCASDAN 59  
 DB 1 MRVAGIRNRYOHLWKGLN--LGLILMFCVADLWTVIYGVPRKATTLFCASDAK 58

QY 60 LTSTKHNWVASQCVPTDPDPHEYLNTVDFNINWENYVEQMEDIIISLWQSLKPC 119  
 DB 59 SYEPAAHNIWATHACVPTDPDPREIEMENVTFENMKNKNVVEQMHEDIISLWQSLKPC 118

QY 120 IOMTFMCIOMNCTDKNNNTSGTNRSSSE-----NPMKTCENITTLVKDKKKEKQAL 174  
 DB 119 VKLTPCLVTLNAGNKTNGNTTNGEOMMEKGMKCSFNITTVISDKKKQVHAL 178

QY 175 FYVSDLAKLADNNTNM-----YTLNCNHTIKQACPKVSEPIPIYCAAGVAIF 228  
 DB 179 FYRLOVVPIDDDNSANTSNWYRLNLCNCSAITQACPKVTEPIPIHYCAPAGFAIL 238

QY 229 KCNSAEFNGTKCNSISVVTCTHGIKPTVSTQLIINGTLSEKIRIMKNSDSGKNIIY 288  
 DB 239 KCDKKNGTGPKCKSVTVQCTHGRPVVSTQLLNGSLAEELIIRSENITNNVKIIV 298

QY 289 TLSSDIEITCVRPNGNQTVEKMGIPMAWYSMALGTGNSRSRV-----AYCQYNTWE 342  
 DB 299 HLNESVEINCRPNKTRQSTPIG-----LQALYTTTRIKGDIRQAYCNISAAWN 350

QY 343 KALANTAERYLELNNTEGNTMIFNRSQDSQDVETHLHFNCHGEFFCYNTSEMFYTF 402  
 DB 351 KTLQOVAKKGLDLN-----QTTIIF-KPPAGGDPETHTSHSFCGGEFFCYNTSLFNSTW 405

QY 403 LCNCTNCTNOSINSANGM-IPCKLKQVVRWSRGGSLYAPPIPGNLTCTSHITGMILQ 461  
 DB 406 --NSTWN--DTLNSGKIKLPCRIKQIINNMGVGRVAPPIEGLKCTNSITGLLT 462

QY 462 MDAPWNKTEN-TFRPIGDMKDIWRNELFYKYVVRVPFVSVAFTPIAPVIGTGTREKR 520  
 DB 463 RDGVNNSSTNETFRPGGDMKDNRNELYKYVVRVIEPIPLGIAPTRAKRRVY---EREK 518

QY 521 AVGLGMLFLGVLNAGSTMGAAATALTQVSHVYKIGVQQDNLRLAIQAQQELLRLSVW 580  
 DB 519 AIGLGAVFLGAGSTMGAVSALTQCARQLLSGIVQQQNLRLRAIEAQHMLQTLTW 578

QY 581 GROLRLARLALAEFLIQOQLLNWCGKRLICYTSYKWNETHWNTNINOINWNLWQ 640  
 DB 579 GIKLOARLAVESYLKQDQLLGIWCGSGRICTTTPVWSSWSN-KSLBEIWNMTWIE 637

QY 641 WDOQIDNVSTIYEEIOKAQVQOQNEKKLELDEWASLWNLDTKWLWYIKITAIIVG 700  
 DB 638 WEREIDNTGVISLIENSQIQEKNQDQLLQDKWASLWNLWFSITKWLWIKIFIMIVG 697

QY 701 ALIGVRIVMIVNLVRNIRQCYQLSLQIPTRQOSEAETPGRTGEGGDEGRPLIPSPQ 760  
 DB 698 GLIGLRIVFVLSLVNRVQSGYSPSLFTLLPAPRDPBPBGIEEGEGQGRSIRLVN 757

QY 761 GFPLPLTDTURTIIILWSYHLSNLISGTQVISHRLGLWILGOKIIDACRICAAVHYW 820  
 DB 758 GFSALFWDLLRNLCLFSYHRLDLILIAITRIVE-----LLGRGWFAIKYLSLQYW 810

QY 821 LOELQKATSILDTFAVANWTTDIIILGRLGRGILNIPRRVROGFERSLL 873  
 DB 811 TQELKNSFISLNNATAVAEGTDRIELIRRAFRVLHPIRRVROGLERALL 863

RESULT 7  
 ENV\_HV1S1  
 ID ENV\_HV1S1 STANDARD; PRT; 847 AA.  
 AC P19550;  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE  
 GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)]  
 GN ENV.  
 OC Human immunodeficiency virus type 1 (SF162 isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90347835; PubMed-2384920;  
 RA Cheng-Mayer C., Quiroga M., Tung J.W., Dina D., Levy J.;  
 RT "viral determinants of human immunodeficiency virus type 1 T-cell or  
 macrophage tropism, cytopathogenicity, and CD4 antigen modulation.";  
 RL J. Virol. 64:4390-4398(1990).  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M65024; AAA45072.1; -  
 DR HIV; M38428; ENVSSP162.  
 DR INTERPRO; IPR000228; -  
 DR INTERPRO; IPR000777; -  
 DR PFAM; PF00516; GP120; 1.  
 DR PFAM; PF00517; GP41; 1.  
 KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;  
 KW Signal.  
 FT SIGNAL 1 29  
 FT CHAIN 30 502 EXTERIOR MEMBRANE GLYCOPROTEIN.  
 FT CHAIN 503 847 TRANSMEMBRANE GLYCOPROTEIN.  
 FT DISULFID 53 73 BY SIMILARITY.  
 FT DISULFID 118 203 BY SIMILARITY.  
 FT DISULFID 125 194 BY SIMILARITY.  
 FT DISULFID 130 155 BY SIMILARITY.  
 FT DISULFID 216 245 BY SIMILARITY.  
 FT DISULFID 226 237 BY SIMILARITY.  
 FT DISULFID 294 328 BY SIMILARITY.  
 FT DISULFID 374 435 BY SIMILARITY.  
 FT DISULFID 381 408 BY SIMILARITY.  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 232 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 274 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 329 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 382 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 388 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 401 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 438 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 454 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 607 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 628 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 847 AA; 96135 MW; 0A901317FD7F2AB CRC64;

Query Match 48.9%; Score 2259; DB 1; Length 847;  
 Best Local Similarity 49.9%; Pred. No. 2.2e-156;  
 Matches 439; Conservative 155; Mismatches 243; Indels 42; Gaps 15;

QY 5 MRAMGRNR-----KLGILYIVMALIIPCLSSQLYATYAGVPVWEDAAAPVLCASDAN 59  
 DB 1 MRVKGIRKYNQHLWRGGTLLGLMLI--CSAVEKLWVYVYGVVWKEATTLFCASDAK 58

QY 60 LSTSTKHNVWASQACVPTDPTPEHVLNNTVDNFNINWYVQEQEDIIISLWDQSLKPC 119  
 DB 59 AYDTEVHNWATHACVPTDPNPQETVLENVTENFMNKNMNVQEQHEDIIISLWDQSLKPC 118

QY 120 IQMTFCWCMCTDKNNNTSTENRTSSSENPMKTECFNITTVLKKKKQKQALFVSD 179  
 DB 119 VKLTPCVTLHCTNLKATNTKSSNWKEMDGEIKNSCFVTTTSIRNKQKAYALFKLD 178

QY 180 LTKLADNTTNTWTLINCNTTTIKQAPKVSFPIPYCAPAGYAFICNSAEFNGTG 239  
 DB 179 VVPI-DND--NTSKYLCNCTSVITQACPKVSFPIPIHYCAPAGFALLCKDKKNGSG 235

QY 240 KCSNIVTCHGKPTVSTQILNGLTSKRIKIMKNTSDGKNIIIVLSSDIEITCV 299  
 DB 236 PCTNVSTVQCTHGRPVYSTQLLNGSLAEGVVRSENFTDNKTIIVQLKESVEINCT 295

QY 300 RPNQNTVOEMKIGP-WANYSMALGTGNSRVAICYNTTEWEKALKNTAERYLELNN 358  
 DB 296 RP-NNNTRKSTITGPGAFY--ATGDIIGDIRQAHCHNISEKW---NNTLQIVTKLQA 348

QY 359 TEGNTTIFNRSDGSDVEYTHLHFNCHGEFFVCNTSEMNTYFLCNGTNCNNTQINS 418  
 DB 349 QFGNKTIVFQKS--SGGDEPILVMSFNGGEGFFVCNTSOLFNSW-----NNTIGPNNT 400

QY 419 NGMI--PCKLKQVVRWVRGSGGLYAPPIPNLTCTSHITGMQMDA--PNKNTENFR 474  
 DB 401 NGTITLPCRIKQINRQEVGKAMYPPIRGQIRCSNITGLLLTRDGEKETSNTETFR 460

QY 475 PIGDMKDIWRNELFKYKVVVPFVAPPTIARVIGTGHREKRAVGLMFLGVLISA 534  
 DB 461 PGGDMKDNRRSELYKYKVVVPLGAVPTAKRRVV---QREKRAVTLGAMFLGFLGA 516

QY 535 AGSTMGAATATLAVQVTHSVKIGVQOQDNLRAIQAOQELLRLSVWGLRQLPARLALET 594  
 DB 517 AGSTMGARSULTVQARQLLSGIVQOQNLRAIEAQHLLQLTVWGLKQARVAYER 576

QY 595 LIQOQLLNLWCKGLICYTSVKNNETWRNTNINQWGLNLTQEWQOQIDNVSSTIYE 654  
 DB 577 YLQOQLLNLWCKGLICYTSVKNNETWRNTNINQWGLNLTQEWQOQIDNVSSTIYE 635

QY 655 ELQKAOVQOQEQNEKKLELDEWASLWNLDTIKLWYIKIALLIIVGALIGVRIWVNLN 714

DB 636 LIEESQOQNEQELLELDKQASLWNNFDSKWLWYKIFIMVGLVGLRIVFTVLSI 695  
 QY 715 VRNIRQGVQPLSLQIPTRQOSEATPORTGEGGDEGRPLIPSPQGFLLYTLDTLRTII 774  
 DB 696 VNRVQGYSPLSQTRFPAPRPGDRPEGIEEGGERDRSSPLVHGLLALIWDLRLSLC 755  
 QY 775 LMSYHLLSNLISGTQTQVISHLRLGLWLTGKIIDACRICAAVHYWLOELQKQATSIDT 834  
 DB 756 LFSYHRLDLILTAARIVE-----LLGRGWALKYGNLQYWIOLKNSAVSLFDA 808

QY 835 FAVAVANWTDIILGILQRLGRILNIPRRVQGFERSLL 873  
 DB 809 IATAVAEGTDRIIEVAQRIAGRAFLHIPRRIRQGFERALL 847

RESULT 8  
 ENV\_HVLC4  
 ID ENV\_HVLC4 STANDARD; PRT; 868 AA.  
 AC P05879;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE  
 DE GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].  
 GN ENV.  
 OS Human Immunodeficiency virus type 1 (CDC-451 isolate) (HIV-1).  
 OC Viruses; Retrov. viruses; Retroviridae; Lentivirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-87041461; PubMed-3490666;  
 RA Desai S.M., Kalyanaram V.S., Casey J.M., Srinivasan A.,  
 RA Andersen P.R., Devare S.G.;  
 RT "Molecular cloning and primary nucleotide sequence analysis of a  
 RT distinct human immunodeficiency virus isolate reveal significant  
 RT divergence in its genomic sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8380-8384(1986).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; M13137; AAA44311.1;  
 DR PIR; C25523; VCLJH4.  
 DR HIV; M13137; ENVSCDC45.  
 DR INTERPRO; IPR000328;  
 DR INTERPRO; IPR00077;  
 DR PFAM; PF00516; GP120; 1.  
 DR PFAM; PF00517; GP41; 1.  
 KW AIDS; Coat protein; Polypeptide; Glycoprotein; Transmembrane;  
 KW Signal.  
 FT SIGNAL 1 32  
 FT CHAIN 33 522 EXTERIOR MEMBRANE GLYCOPROTEIN.  
 FT CHAIN 523 868 TRANSMEMBRANE GLYCOPROTEIN.  
 FT FT 55 75 BY SIMILARITY.  
 FT DISULFID 120 216 BY SIMILARITY.  
 FT DISULFID 127 207 BY SIMILARITY.  
 FT DISULFID 132 163 BY SIMILARITY.  
 FT DISULFID 229 258 BY SIMILARITY.  
 FT DISULFID 239 250 BY SIMILARITY.  
 FT DISULFID 307 341 BY SIMILARITY.  
 FT DISULFID 387 456 BY SIMILARITY.  
 FT DISULFID 394 429 BY SIMILARITY.  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 342 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 395 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 828 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 868 AA; 98698 MW; A11527FC52A6F0C8 CRC64;

Query Match 48.68; Score 2247.5; DB 1; Length 868;
Best Local Similarity 49.28; Pred. No. 1.6e-155;
Matches 439; Conservative 159; Mismatches 249; Indels 45; Gaps 18;

QY 3 VTMRAGKRRNR-----KLGILYIYVWALICLSSQLYATVYAGVPVEDAAPVFCASD 57
DB 1 MAMRAGKRRNCQLRWGTMGLMLMI--CSAANLVTVYGVYVWKEATTLFCASD 58

QY 58 ANLTSTKHNVASQACVPDPTPHEYLNTVDFNFWENYVQMOEDIIISLWDQSLK 117
DB AKAYDTEAHNVWACHVPTNPVQVLENVFNFNKNNVWVQMHEDIIISLWDQSLK 118

QY 118 PCIQMTFMCIONCTDIKNNTSTENRT-----SSSENPMKTCERNITVLKDKREK 171
DB 119 PCVKLTPLCVTLNCTDNTNTTTELSIIVWQKGMKMRNCSNITTSIDRKVORE 178

QY 172 QALPYVSDAKLADN--NTMTTLNCSNTTIKQCPKVSFPIPIYCAPAGYAEK 229
DB 179 YALPYKLDVPEIDNKNNTTKYLLNCSNTSVITQCPKVSFPIPIHYCTPGFALLK 238

QY 230 CNSAEFNGTGKCSNISVVTCTHGIKPTVSTQILNGLSKIRIMKNISDSGKNIYV 289
DB 239 CNDKKFNGTGCTNVSTVQCTHGIRPVVSTQILLNGSLAEVWIRSENFNTNAKTIIV 298

QY 290 LSSDIEITCVPRGNQIVQEMKIGP--MAWYSMALGTGS--NRSRVAYCOYNTTEWAKLK 346
DB 299 LNVSEVLEINCRTP--NHHTKRRTLTGPRVWYV----TGEILGNIRQAHCNISRAQW---N 349

QY 347 NTABRYLELINTEGNTMTFNRQDGSDEVEVTHLHCHGEFFYCNTESEFN--YVFLC 404
DB 350 NTLQQAATTLREQGNTIAFNQS--SGDDPEIVMHSFNCGGEFFYCNSTQFNSANWVTS 408

QY 405 NGT--NCNNTQSTANGMIPCKLQVYRSMWGGSLYAPPICPNLTCISHITGMILQMD 463
DB 409 NGTVSVTRKQDGTIIITLPCRIQIINRWQVGVKMYALPIKGLIRCSSNITGLLTRD 468

QY 464 APW--NKTENFRPFGGDKMDIWRNELKFKYVVRKVPSPVAPTPIARPVIGTTHREKRV 522
DB 469 GGGENQTEIFRPGGGMDNRWSELYKYVVKIEPLGVAPTAKRRVW----QREKRV 524

QY 523 G--LGMLFLGVLSAAGTMAAATALTQTHSVKIGVQQQDNLLRATQAOQELLRLSWG 581
DB 525 GMLGAMFLGFLGAGSTMGATSMALTQVQAROLLGSIQQQNNLLRAKAOQHLQLTWWG 584

QY 582 IRLRLARLLALETIQNQLNLWGCKRGLICYTSVKNWNETWRNTTINQIWGNLTWQEW 641

```

```

DB 585 IKQLQARILAVERYLKDQQLLQWGCCKLICCTAVPWNASWSNKT--LDQIWNNTWNEW 643
QY 642 DOQIDNVSSITIEEIOKAQVQOQNEKKLELDEWASIWNNWLDITKWLYIKIILVGA 701
DB 644 DREIDNTHLITIEESQOQEKQOQELQDKWASLWSDITKWLYIKIFIMVGG 703
QY 702 LIGVRIYVILNVRNIROGYPLSLQIPTRQOSEATPGRTGGGDEGRPLIPSPQG 761
DB 704 LIGLRIYFAVLGIYVNRVQGYSPLSFQTLNPNRGPDRPECTEGSGGERGSTRLVHG 763
QY 762 FLPLYLDRTIILWSYHLLSNLSIGTQTVISHLRLGWLILGQIIDACRCAAVIHWL 821
DB 764 FLALVMDLRSCLFSYHRLRLLLIYARIVE-----LLGRGWEVLKYWNLLQYWS 816
QY 822 QELQKATSITDFAVAVANWTDIILGIORLGRGILNIPRVRQGFERSLL 873
DB 817 QELKNSAVSLVNVTAIAVAGTDRVIEVQYIRAFIHPRIIRQGFERALL 868

RESULT 9
ENV_HV1Y2
ID ENV_HV1Y2 STANDARD; PRT; 843 AA.
AC P35961;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE
DE GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].
GN ENV.
OS Human immunodeficiency virus type 1 (YU-2 isolate) (HIV-1).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93021387; PubMed=1404605;
RA Li Y., Hui H., Burgess C.J., Price R.W., Sharp P.M., Hahn B.H.,
RA Shaw G.M.;
RT "Complete nucleotide sequence, genome organization, and biological
RT properties of human immunodeficiency virus type 1 in vivo: evidence
RT for limited defectiveness and complementation.";
RL J. Virol. 66:6587-6600(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M93258; ; NOT_ANNOTATED_CDS.
DR PIR; H44001; H44001.
DR INTERPRO; IPR000328; .
DR INTERPRO; IPR000777; .
DR PFAM; PF00516; GP120; 1.
DR PFAM; PF00517; GP41; 1.
DR AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
DR Signal.
FT SIGNAL 1 29
FT CHAIN 30 489 EXTERIOR MEMBRANE GLYCOPROTEIN.
FT CHAIN 490 843 TRANSMEMBRANE GLYCOPROTEIN.
FT TRANSMEM 738 755 POTENTIAL.
FT DISULFID 53 73 BY SIMILARITY.
FT DISULFID 118 201 BY SIMILARITY.
FT DISULFID 125 192 BY SIMILARITY.
FT DISULFID 130 155 BY SIMILARITY.
FT DISULFID 214 243 BY SIMILARITY.
FT DISULFID 224 235 BY SIMILARITY.
FT DISULFID 292 326 BY SIMILARITY.
FT DISULFID 373 432 BY SIMILARITY.
FT DISULFID 380 405 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).

```



FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 384 384 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 608 608 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 622 622 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 853 AA; 97043 MW; 849B0B8CB8F7008 CRC64;

Query Match 48.5%; Score 2242; DB 1; Length 853;  
 Best Local Similarity 49.9%; Pred. No. 3.8e-155;  
 Matches 442; Conservative 150; Mismatches 245; Indels 48; Gaps 15;

QY 5 MRANG-KRN-----RKLGLYIVMALIIPCLSSOLYATYAGVPVWEDAAPVLCASDAN 59  
 DB 1 MRVGRICRNCNLKWKGLM--LLGLMTCSNADNLWTVYGVVWKEATTLFCASDAK 58  
 QY 60 LTSTKHNWASQACVTPDPPEHYLLTNVTDNFNINWYVQWQEDIIISLWQSLKPC 119  
 DB 59 SYKTEAHNIWATHACVTPDPPEHYLLTNVTDNFNINWYVQWQEDIIISLWQSLKPC 118  
 QY 120 IQMTFMCIQMCTD-----IKNNSTSGTENRTSSSENPMKTCFNTTTLVKDKKKQ 172  
 DB 119 VKLPVLCVTLNCIDEVMTMKNNV-----TEIRMKNSFNITTVVRDKTKQVH 170  
 QY 173 ALFYVSDLTKLADNNTN-TMYTLINCNTTIKQACPKVSEPIPIYCAPAGYAIFKCN 231  
 DB 171 ALFYRLDIVPIDNDNSTNTYRLNCNTSAITQACPKVSEPIPIHYCAPAGFAILKCR 230  
 QY 232 SAENGTKGCSNIVVTGTHGKIVTSTOLLNLTLSKEKIRINGKNISDSGKNIIVTLS 291  
 DB 231 DKRFGTGPCTNVSTVQCTHGIRPVSTQLLNGSLAEELIIRSENLTNAKIIIVQLN 290  
 QY 292 SDIEITVRPNNQTVQEMKIGPMWYSMALGTGNSRSV--AYCOYNTTWEKALNTA 349  
 DB 291 ESVAINTTRPVN-IRQTSIG---LQALYTKTSIIQOALCNISKNWNKTLQOVA 345  
 QY 350 ERYLELINTEGNTMIFNRSQDGSDEVTHLHFNCHGEFFYCNTESEMYTFLCNGTNC 409  
 DB 346 IKLGNLLN---KTTIIFKPS-SGGDEPHTHSPNGGEPFYCNTESEMYTFLCNGTNC 400  
 QY 410 NNTOSINSANGMPCIKLQOVVSRWMRGSGLYAPPIGNLTICSHITGMILQDAPNKT 469  
 DB 401 ANSTESDPPKTLQCRKIQIINMNOGVKAMAPPYIEQINCSSNITGLLTRDGGVNS 460  
 QY 470 EN-TFRIGGDMKDIWELNFKYKVRKVPFVAPTPTARPVIGTGHREKRAVLGLMLF 528  
 DB 461 SNETFRGGDMRNRWSELYKYKVKIEPLGVAPTAKRVV-----ERKRAVLGLMLF 516  
 QY 529 LGVLSAAGSTMGAAATLTQVTHSVIKGIVQOQDNLRAIQAOQELRLSVWGIQRLAR 588  
 DB 517 LGFLGAAGSTMGARSLLTQVQARLLSGIVQOQDNLRAIQAOQELRLSVWGIQRLAR 576  
 QY 589 LLALETIQOQLNLNCGCKRGLICYSVKKNWTRWNTNINQIWNLTWQEDQOQIDNV 648  
 DB 577 ILAVERIKDQQLGIGCGSKLICCTTTPVWNSWSN-RSLNDIWNQNTWMEWEREIDNY 635

QY 649 SSTIYEIQAKVQOQNEKRLLEDEWASLWNLWDITKWLWYIKIAIIVGALIGVRIV 708  
 DB 636 TGLYRLIEESQTQOERNEQELLEDKWASLWNLWNTQWLWYIKIFIMVGLIGLRIV 695  
 QY 709 MIVNLVNRQGYQPLSLQIPTRQQAETPGRTGGGDEGRPLIPSPQGFPLPLYT 768  
 DB 696 FAVLSLVNRVQGYSPSLFQTLPPAPRGDPDPEEGEGGERGSRIRLVNGFSALWD 755  
 QY 769 DLRTIILWSVHLLSNLSGTQTVISHRLGLWILGOKIIDACRIACAAVHVLWLOLQKSA 828  
 DB 756 DLRLNCLFSYHRLDLILIAARIVE-----LLGRGWELKYLWNLQWISRELKNSA 808  
 QY 829 TSLDITFAVAVANWTDDIILGIGRLGILNIPRRVQGFERSLL 873  
 DB 809 SSLDLTIAIAVAECTDRVIEVRRACRAVLHPIRQGLERLLL 853

RESULT 11  
 ENV\_RV153  
 ID ENV\_RV153 STANDARD; PRT; 852 AA.  
 AC P19549;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE ENVELOPE POLYPROTEIN GP160 PRECURSOR [CONTAINS: EXTERIOR MEMBRANE  
 DE GLYCOPROTEIN (GP120); TRANSMEMBRANE GLYCOPROTEIN (GP41)].  
 GN ENV.  
 OS Human immunodeficiency virus type 1 (SF33 isolate) (HIV-1).  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 RN [1]  
 RP MEDLINE-90317906; PubMed-2370688;  
 RA York-Higgins D., Cheng-Mayer C., Bauer D., Levy J.A., Dina D.;  
 RT "Human immunodeficiency virus type 1 cellular host range,  
 RT replication, and cytopathicity are linked to the envelope region of  
 RT the viral genome.";  
 J. Virol. 64:4016-4020(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M38427; AAA45067.1; -;  
 DR HIV; M38427; ENV5SF33.  
 DR INTERPRO; IPR000328; -;  
 DR INTERPRO; IPR000777; -;  
 DR PFAM; PF00516; GP120; 1.  
 DR PFAM; PF00517; GP41; 1.  
 KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;  
 KW Signal.  
 FT CHAIN 1 29  
 FT CHAIN 30 506 EXTERIOR MEMBRANE GLYCOPROTEIN.  
 FT CHAIN 507 852 TRANSMEMBRANE GLYCOPROTEIN.  
 FT DISULFID 53 73 BY SIMILARITY.  
 FT DISULFID 118 206 BY SIMILARITY.  
 FT DISULFID 125 197 BY SIMILARITY.  
 FT DISULFID 130 156 BY SIMILARITY.  
 FT DISULFID 219 248 BY SIMILARITY.  
 FT DISULFID 229 240 BY SIMILARITY.  
 FT DISULFID 297 331 BY SIMILARITY.  
 FT DISULFID 377 439 BY SIMILARITY.  
 FT DISULFID 384 412 BY SIMILARITY.  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).





[illegible]





FT CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 384 384 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 442 442 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 612 612 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 812 812 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 852 AA; 97203 MW; 2BB866345DEC915F CRC64;

Query Match 48.2%; Score 2228.5; DB 1; Length 852;  
 Best Local Similarity 49.4%; Pred. No. 3.6e-154;  
 Matches 442; Conservative 153; Mismatches 229; Indels 71; Gaps 20;

QY 5 MRANG-KRNRK-----LGILYIVMALIIPCLSSQLYATYAGVPVWEDAAPVLF 53  
 DB 1 MRVGIKKNYQHLRWGGMGLGIMI-----CSATDKLWTVYVYGVVWKEANTLIF 53

QY 54 CASDANLSTKHNWASQACVTDTPHYLELTNTDNFNWENTNVEOMQEDIIISLWD 113  
 DB 54 CASDAKAYDEIHNWATHACVTPDNPQELVMGNTNFNMWKNDAVQMHEDIISLWD 113

QY 114 QSLKPCIQMTFMCIQNMCTDIKNNTSGTENSTSSSE-NPMKTCFENITVLKDKKEKQ 172  
 DB 114 QSLKPCVKLTPCLVTLNCHDF--NATNATNSGKMEGEMKNCSEFNITTSIRDKMKEY 171

QY 173 ALFYVSDLKLANNTNTMTYLINCNSTTIKQACPKVSFEPIPIYICAPAGYAFKNS 232  
 DB 172 ALFYKLDIVPI-DNDKTNTRYLISNTSVITQACPKVTFEPIPIHYCAPAGFAILKCNN 230

QY 233 AEFNGTKCSNIVSWTCTHGKPTVSTQILNGLTSKEKIRINGKNISSDGNIIIVTSS 292  
 DB 231 KFEKGTGCTNVTSTVCTHGIRPVSTQILNGLTSKEKIRINGKNISSDGNIIIVTSS 290

QY 293 DIBITVRPNNQTOEMKIGP-MAWYSMALGTGNSRVSAYCOYNTTEWEKALKNTAER 351  
 DB 291 SVEINCTRP-NNNTRKRITMGPVYVYTGQIIGDIR--AHCNLSRSKWE-----NTLKQ 343

QY 352 YLELINTSGNTMLFNRSQSDGVETHLHFNCHGEFFYCNTESEMYTFLCNGTNCNN 411  
 DB 344 IVTKLRVQKNTIIVFNRS-SGGDPEIIVMHSFNCGGEFFFCNTOLFNTSTWY-----RNT 397

QY 412 TOSINSANGMI--PCKLKQVVRSMRGGSGLYAPPIGNLTCISHTMILQMDAPWNT 469  
 DB 398 TGNITEGNSPIPLPCRIKQIINWQEVGKAMYAPPICQIKCSSNITGLLTRDG-GNNN 456

QY 470 ENT-----FRPIGGDKMDWRNELFKYKVVVVKPVSFAPTPARFVIGTGRKRRVGLG 525  
 DB 457 ETTDEIFRPGGNNRDNWRSELYKYKVVVVKIEPLGVAPTAKRRV---QREKRAVGLG 512

QY 526 MLFLGLVSAAGSTMGAAATALTVOHSHVYKIGVQOQDNLPAIOAQOELLRLSVGIRQL 585  
 DB 513 ALFLGLGAAGSTMGAAATALTVOHSHVYKIGVQOQDNLPAIOAQOELLRLSVGIRQL 572

QY 586 RARLLAETLTONQOLLNMGCKRLCYTSVKWNTRNTNINQWGLNTWOEWDQOI 645  
 DB 573 QARVLAVERYLKQOLLGICGKLICTTAVPWNASWSN-KSLSDIWDNNTWMEWEREI 631

QY 646 DNVSTIYEETQKAOVQOQEKLEDEWASLWNLDTKWLWYKIAIIVIGALIGV 705  
 DB 632 DNYTNLIYSLIEDSQIQEKNEKLEDEWASLWNLDTKWLWYKIAIIVIGALIGV 691

QY 706 RIVMIVLNVRNIROGYQPLSLQIPTROQSEAEETPRTG-----EGGDEGRPLIP 757  
 DB 692 RIVFAVLSIVNRVROGYSPLSFO-----TRLFGRGPDPRPEIEEGGERDRDRSSP 743

QY 758 SPQGFPLIYTDRLRIIILWSYHLLSNLISGTTQTVISHRLGLIWLKIIIDACRICAAVI 817  
 DB 744 LYDGFALFWDLRSFLFSYHRLRDLIIIVTRIVE-----LLGRGWEVLKYWNLL 796

QY 818 HYWLQELOKSATSLIDTFAVAVANNWTDIIILGILQRLGRGILNIPRVQGFERSL 872  
 DB 797 QIWSQELKNSAVSLUNATAIAGERTDRAIEVVOFAFRAILHPRIRIROGLERAL 851

Search completed: January 25, 2001, 11:02:20  
 Job time: 655 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2001, 10:51:24 ; Search time 15.24 Seconds  
(without alignments)  
1028.642 Million cell updates/sec

Title: US-08-911-824-61  
Perfect score: 4622  
Sequence: 1 MIVTMRANGKRNKLGILYI.....GKILNIPRRVQGFERSLL 873

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgn2\_6/prodata/2/1aa/3A\_COMB.pep.\*
- 2: /cgn2\_6/prodata/2/1aa/3B\_COMB.pep.\*
- 3: /cgn2\_6/prodata/2/1aa/6\_COMB.pep.\*
- 4: /cgn2\_6/prodata/2/1aa/PCUS\_COMB.pep.\*
- 5: /cgn2\_6/prodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4622	100.0	873	2	US-08-912-129A-61
2	2250	48.7	855	1	US-08-022-835-6
3	2250	48.7	855	1	US-08-388-809-6
4	2250	48.7	855	2	US-08-647-714-6
5	2248	48.6	857	2	US-08-448-603A-30
6	2248	48.6	857	3	US-09-134-075-30
7	2247	48.6	850	2	US-08-448-603A-28
8	2247	48.6	850	3	US-09-134-075-28
9	2211.5	47.8	856	1	US-08-022-835-2
10	2211.5	47.8	856	1	US-08-388-809-2
11	2211.5	47.8	856	2	US-08-647-714-2
12	2211.5	47.8	863	3	US-08-463-210-11
13	2211	47.8	857	1	US-08-022-835-4
14	2211	47.8	857	1	US-08-388-809-4
15	2211	47.8	857	2	US-08-647-714-4
16	2208.5	47.8	856	2	US-07-916-098A-2
17	2205.5	47.7	861	1	US-08-127-499A-14
18	2205.5	47.7	861	1	US-08-482-847-14
19	2190.5	47.4	880	3	US-08-788-815-7
20	2190.5	47.4	880	3	US-09-157-963-7
21	2181.5	47.2	826	1	US-08-375-510-2
22	2181.5	47.2	826	1	US-08-487-657-2
23	2175.5	47.1	856	1	US-08-375-100-1
24	1926.5	41.7	618	2	US-08-912-129A-54
25	1921	41.6	373	2	US-08-912-129A-52
26	1791.5	38.8	665	2	US-08-448-603A-32
27	1791.5	38.8	665	3	US-09-134-075-32
28	1682	36.4	600	2	US-08-392-806A-4

29	1678	36.3	606	2	US-08-392-806A-6
30	1657	35.9	596	2	US-08-392-806A-2
31	1452	31.4	519	1	US-08-589-446-8
32	1452	31.4	519	1	US-08-444-882-8
33	1452	31.4	519	2	US-08-389-459A-8
34	1452	31.4	519	3	US-08-987-867A-8
35	1428.5	30.9	351	1	US-08-470-202-46
36	1428.5	30.9	351	1	US-08-471-770-46
37	1428.5	30.9	351	2	US-08-468-059-46
38	1346	29.1	356	1	US-08-602-713-12
39	1346	29.1	356	3	US-08-989-493-12
40	1316	28.5	494	3	US-08-889-841-19
41	1314.5	28.4	495	3	US-08-889-841-28
42	1308.5	28.3	474	3	US-08-889-841-36
43	1308.5	28.3	474	3	US-08-889-841-39
44	1308.5	28.3	495	3	US-08-889-841-25
45	1305.5	28.2	486	3	US-08-889-841-8

ALIGNMENTS

RESULT 1  
US-08-912-129A-61  
; Sequence 61, Application US/08912129A  
; Patent No 5922533  
; GENERAL INFORMATION:  
; APPLICANT: VALLARI, ANADRUZELA S.  
; APPLICANT: HACKITT, JOHN JR.  
; APPLICANT: HICKMAN, ROBERT K.  
; APPLICANT: VARIATEK, VINCENT A. JR.  
; APPLICANT: NECKLAWS, ELIZABETH A.  
; APPLICANT: GOLDEN, ALAN M.  
; APPLICANT: BRENNAN, CATHERINE A.  
; APPLICANT: DEWARE, SUSHIL G.  
; TITLE OF INVENTION: RAPID ASSAY FOR SIMULTANEOUS DETECTION AND DIFFERENTIATION

; NUMBER OF SEQUENCES: 89  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette, 1.44 MB  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS (Windows 95)  
; SOFTWARE: Microsoft Word (ASCII format output)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/912,129A  
; FILING DATE: 15-AUG-1997  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Danckerts, Andreas M.  
; REGISTRATION NUMBER: 32,652  
; REFERENCE/DOCKET NUMBER: 6109.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847-937-9803  
; TELEFAX: 847-538-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 61:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 873 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; US-08-912-129A-61

```

Query Match      100.0%; Score 4622; DB 2; Length 873;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYTMRAMGRNKLGLIYVMAIIICLSSSOLYATYVAGVPVWEDAAAPVLFCSADANL 60
DB 1 MYTMRAMGRNKLGLIYVMAIIICLSSSOLYATYVAGVPVWEDAAAPVLFCSADANL 60
QY 61 TSTKHNWASQACVPTDPTPHEYLNTVNTDNFNINWNTWVQMQEDIIISLWQSLKPCI 120
DB 61 TSTKHNWASQACVPTDPTPHEYLNTVNTDNFNINWNTWVQMQEDIIISLWQSLKPCI 120
QY 121 QMTFMCIONMCTDIKNNTSGTENTSSSENPMKTCFNTITVLDKKEKQALFYVSD 180
DB 121 QMTFMCIONMCTDIKNNTSGTENTSSSENPMKTCFNTITVLDKKEKQALFYVSD 180
QY 181 TKLADNNTNTMTYLNCNSTIKQACPKVSEPIPIYCAPAGYAIFKCSAEFNGTK 240
DB 181 TKLADNNTNTMTYLNCNSTIKQACPKVSEPIPIYCAPAGYAIFKCSAEFNGTK 240
QY 241 CSNISVVTCTHGIKPTVSTQILNGLTSLKEKIRIMGNISDGNKIIVTLSSDIEITCVR 300
DB 241 CSNISVVTCTHGIKPTVSTQILNGLTSLKEKIRIMGNISDGNKIIVTLSSDIEITCVR 300
QY 301 PGNNQTVQEMKIGPMWNTSMALGTGNSRNVAYCOYNTTEWEKALNTAERYLELNINTE 360
DB 301 PGNNQTVQEMKIGPMWNTSMALGTGNSRNVAYCOYNTTEWEKALNTAERYLELNINTE 360
QY 361 GNTTIFNRSQDGSVEVTHLFHNCHEFFCYNTSEMNTFLCNGTNCNNTQINSANG 420
DB 361 GNTTIFNRSQDGSVEVTHLFHNCHEFFCYNTSEMNTFLCNGTNCNNTQINSANG 420
QY 421 MIPCKLKQVVRSMRGGSLYAPPIGNLTCSHITGMILQMDAPWKNKTENTFRPIGDM 480
DB 421 MIPCKLKQVVRSMRGGSLYAPPIGNLTCSHITGMILQMDAPWKNKTENTFRPIGDM 480
QY 481 KDINWELFKYVVRKVPFVAPPIAPVPIGTGTHREKRAVGLGMLFLGVLAAAGSTM 540
DB 481 KDINWELFKYVVRKVPFVAPPIAPVPIGTGTHREKRAVGLGMLFLGVLAAAGSTM 540
QY 541 AAATALTAVQTHSVIKIVQOQDLRAIQAOQELRLSVGIRQLRALLALETIONQ 600
DB 541 AAATALTAVQTHSVIKIVQOQDLRAIQAOQELRLSVGIRQLRALLALETIONQ 600
QY 601 LNLWCKGRGLICTSVKKNWETHNTNTNQINQNLFWQEDWQOQIDNVSTIYEEIQAK 660
DB 601 LNLWCKGRGLICTSVKKNWETHNTNTNQINQNLFWQEDWQOQIDNVSTIYEEIQAK 660
QY 661 VQOBNKLELDEWASLWNLIDITKWLVIKIAIIVGALIGRIVMVLNVRNIQ 720
DB 661 VQOBNKLELDEWASLWNLIDITKWLVIKIAIIVGALIGRIVMVLNVRNIQ 720
QY 721 GYQPLSQIPTROQSEATPGRTGEGGDBGRPLIPSPQGFPLLYTLDRITILWSYHL 780
DB 721 GYQPLSQIPTROQSEATPGRTGEGGDBGRPLIPSPQGFPLLYTLDRITILWSYHL 780
QY 781 LSNLSGTQVISHRLGLWILGOKIIDACRICAAVHVLWLOKASATSLDTFAVAVA 840
DB 781 LSNLSGTQVISHRLGLWILGOKIIDACRICAAVHVLWLOKASATSLDTFAVAVA 840
QY 841 NWTDDIILQIRGLRGILNIPRRVQGFERSLL 873
DB 841 NWTDDIILQIRGLRGILNIPRRVQGFERSLL 873

```

RESULT 2

US-08-022-835-6  
; Sequence 6, Application US/08022835  
; Patent No. 5420030  
; GENERAL INFORMATION:  
; APPLICANT: Reitz Jr., Marvin S.  
; APPLICANT: Franchini, Genoveffa

```

APPLICANT: Markham, Phillip D.
APPLICANT: Gallo, Robert C.
APPLICANT: Loxi, Franco C.
APPLICANT: Popovic, Mikulas
APPLICANT: Garter, Suzanne
TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: Eleventh Floor, 1615 L. Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Efloppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/022,835
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/599,491
FILING DATE: 17-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Scott, Watson T.
REGISTRATION NUMBER: 26,581
REFERENCE/DOCKET NUMBER: WTS/5683/78507/SRL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 855 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-022-835-6

```

Query Match 48.7%; Score 2250; DB 1; Length 855;  
Best Local Similarity 51.6%; Pred. No. 2e-176;  
Matches 448; Conservative 138; Mismatches 239; Indels 44; Gaps 18;

```

QY 15 LGILYIVMALIIPCLSSSOLYATYVAGVPVWEDAAAPVLFCSADANLSTKHNWASQAC 74
DB 21 LGILMI-----CNAEKLWTVYVYGVVWKEATVTLFCASDRKAYDTEVHNWATHAC 73
QY 75 VPTDPTPHEYLNTVNTDNFNINWNTWVQMQEDIIISLWQSLKPCIOMTFCIONMCTDI 134
DB 74 VPTDPTPHEYLNTVNTDNFNINWNTWVQMQEDIIISLWQSLKPCIOMTFCIONMCTDI 133
QY 135 KNNNTSGTENTSSSE-----NPMKTCFNTITVLDKKEKQALFYVSDTLKLANNTT 189
DB 134 RNAINGNDTNTSSSRGWSGGEKNCSPNTTINRKGKQYALFYKLDIAPIDNN-S 191
QY 190 NMTYLNCNSTIKQACPKVSEPIPIYCAPAGYAIFKCSAEFNGTKCSNISVVTCT 249
DB 192 NNRYLISCTSVITQACPKVSEPIPIYCAPAGYAIFKCSAEFNGTKCSNISVVTCT 251
QY 250 THGKPTVSTQILNGLTSLKEKIRIMGNISDGNKIIVTLSSDIEITCVRGNNNTVQ 309
DB 252 THGIRPVSTQILNGLTSLKEKIRIMGNISDGNKIIVTLSSDIEITCVRGNNNTVQ 310
QY 310 MKTIGP-MANYSMALGTGNSRNVAYCOYNTTEWEKALNTAERYLELNINTEGNTTIFN 368
DB 311 IHIGPGRVFTTGEIIGDIRQ--AHCNLSRAKWNDFLNKIVIKLREQF---GNKTIVFK 364
QY 369 RSDGSDVEVTHLFHNCHEFFCYNTSEMNTFLCNGTNCNNTQINSANGMIPCKLK 427

```

```

; 365 HSSGDDPEIVTH-SFNCGGDEFFCNSQLENSTW--NVTEESNNVTENNIT--LPCRIK 419
;
; 428 QVSRWMMRGSGLYAPPPIGNLTICSHITGMILQMDA--PWNKTEFTFRIGGDMKDIWR 485
; 420 QIINMQEVRGMYAPPIRQIRCSNITGLLTRDGGPEDNKE-VFRPGGDMRDNWR 478
;
; 486 NELFKYKVRVKPSPVAPTPIARPVIGTGTREKRAVGLGMLFLGVLSSAAGSTMGAAATA 545
; 479 SELYKYKVKVIEPLGVAPTAKRRVV---QREKRAVGIGAVFLGFLGAAGSTMGAAAT 534
;
; 546 LTQVTHSVKIGVQQDNLRAIQAOQELLRLSWGIRQLRALLALETILQOQLNLW 605
; 535 LTQVQARLLSGVQQDNLRAIQAOQELLRLSWGIRQLRALLALETILQOQLNLW 594
;
; 606 GCKRGLICTSVKWNWNTNINQWGLTWOEDQIDNVSSITIEEIOKAQVOOEQ 665
; 595 GCSGKLICTAVPWNASWSN-KSLNKIWNMTWIEWDRNNTYTSIYSLIESQNOQEK 653
;
; 666 NEKKLELDEWASLWNWLDITKWLWYKIAIIVGALIGVRIYVIMVLNVRNRQGYOPL 725
; 654 NEQELLELDKWSLWNWLDITKWLWYKIAIIVGALIGVRIYVIMVLNVRNRQGYOPL 713
;
; 726 SLQIPTRQOSEAETPGTGGGDEGRPRPLIPSPQGFPLLYTDLRTIILWSHLLSNLI 785
; 714 SFQTHLPSSRGPDRPGGIEEGGERDRSGPLVNGFLALIWVDRSLFLSFYHRLRDL 773
;
; 786 SGTQTVISHURL-GLWILGOKIDACRICAAYHYWLOELQKSAATSLIDTFAVANWTD 844
; 774 LIVNRIVELLGLAGW-----EVLKYNWLLQYWSQELKNSAVSLINATAVABGTD 826
;
; 845 DIILGIQRLGRGILNIPRRVROGFERSLL 873
;
; 827 RVIEVLQAVRAILLHPRRQGLERALL 855
;

```

RESULT 3

```

US-08-388-809-6
; Sequence 6, Application US/08388809
; Patent No. 5576000
;
; GENERAL INFORMATION:
; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,
; APPLICANT: GENOVEFFA, MARKHAM, PHILIP D. GALLO, ROBERT
; APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
; APPLICANT: GARTNER, SUZANNE
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK, 3.5"
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/388,809
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,835
; FILING DATE: 25-FEB-1993
; APPLICATION NUMBER: US 07/599,491
; FILING DATE: 17-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: LESLIE A. SERUNIAN
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4092052
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 855 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-388-809-6

Query Match      48.7%; Score 2250; DB 1; Length 855;
Best Local Similarity 51.6%; Pred. No. 2e-176;
Matches 448; Conservative 138; Mismatches 239; Indels 44; Gaps 18;

QY 15 LGILYIVMALIIFCLSSQLYAVYAGVPWEDAAAPVLFCSADANLTSTKHNWASQAC 74
DB 21 LGILMj-----CNAEEKLWVTYVGVVWKEATTTLFCASDRKAYDTEHNVWATHAC 73
QY 75 VPTDPTPHEYLNTDNFNINWYVQOEDIISLWQSLKPCIQMTFCIQMNCIDI 134
DB 74 VPTDPTPHEYLNTDNFNINWYVQOEDIISLWQSLKPCIQMTFCIQMNCIDI 133
QY 135 KNNISGTEENRTSSSE-----NPMKTECFNITTVLKDKKKEKQALFYVSDLTKLADNNT 189
DB 134 RNATNGNDTNTSSRGVGGGEMKNCSEFNITNIRGKVQKEVALFYKLDIAPIDNN-S 191
QY 190 NMYTLLNCSNTIKOACPKVSEPIPIYCAPAGYAFKCNSEAFNGTKCSNISVVC 249
DB 192 NNRYLISCNTSVITQACPKVSEPIPIHYCAPAGFAILCKDKKFGKPGCTNVSTVQC 251
QY 250 THGIKPTVSTQILNGTSLSEKIRIMKNSDSGKNIIVTLSSDIETTCVRPGNNQTVQE 309
DB 252 THGIRPVSTQILLNGSLAEVEVIRSANFADNAKVIIVQLNESVEINCTRP-NNNTRKS 310
QY 310 MKIGP-MANYSMALGTGNSNRVAYCOYNTTEWEKALKNTAERYLELINNTEGNTMIFN 368
DB 311 IHIGPGRAFYTTEIGIDIRQ---AHCNLSRAKWNDTLKNIVIKLREQF---GNKTIVFK 364
QY 369 RSQDSDVEVTHLHFNCHGEFFYCNTSEMYNIFLCNGT-NCNNTQSINSANGMIPCKLK 427
DB 365 HSSGDDPEIVTH-SFNCGGDEFFCNSQLENSTW--NVTEESNNVTENNIT--LPCRIK 419
QY 428 QVSRWMMRGSGLYAPPPIGNLTICSHITGMILQMDA--PWNKTEFTFRIGGDMKDIWR 485
DB 420 QIINMQEVRGMYAPPIRQIRCSNITGLLTRDGGPEDNKE-VFRPGGDMRDNWR 478
QY 486 NELFKYKVRVKPSPVAPTPIARPVIGTGTREKRAVGLGMLFLGVLSSAAGSTMGAAATA 545
DB 479 SELYKYKVKVIEPLGVAPTAKRRVV---QREKRAVGIGAVFLGFLGAAGSTMGAAAT 534
QY 546 LTQVTHSVKIGVQQDNLRAIQAOQELLRLSWGIRQLRALLALETILQOQLNLW 605
DB 535 LTQVQARLLSGVQQDNLRAIQAOQELLRLSWGIRQLRALLALETILQOQLNLW 594
QY 606 GCKRGLICTSVKWNWNTNINQWGLTWOEDQIDNVSSITIEEIOKAQVOOEQ 665
DB 595 GCSGKLICTAVPWNASWSN-KSLNKIWNMTWIEWDRNNTYTSIYSLIESQNOQEK 653
QY 666 NEKKLELDEWASLWNWLDITKWLWYKIAIIVGALIGVRIYVIMVLNVRNRQGYOPL 725
DB 654 NEQELLELDKWSLWNWLDITKWLWYKIAIIVGALIGVRIYVIMVLNVRNRQGYOPL 713
QY 726 SLQIPTRQOSEAETPGTGGGDEGRPRPLIPSPQGFPLLYTDLRTIILWSHLLSNLI 785
DB 714 SFQTHLPSSRGPDRPGGIEEGGERDRSGPLVNGFLALIWVDRSLFLSFYHRLRDL 773
QY 786 SGTQTVISHURL-GLWILGOKIDACRICAAYHYWLOELQKSAATSLIDTFAVANWTD 844
DB 774 LIVNRIVELLGLAGW-----EVLKYNWLLQYWSQELKNSAVSLINATAVABGTD 826
QY 845 DIILGIQRLGRGILNIPRRVROGFERSLL 873

```



Db	696	LAIVFAVLIVNRVQGYSPFLSFQFLPAPRRDPDRPEEGIEESGGEGDRSRLVDGFLA	75
Qy	765	LLYTLRLTILSYHLLSNLISGTQTVISHRLGLWILQKIIDACRIACAIVHYWLQEL	824
Db	756	LIWDDLRLSLCLSYHRLRLDILLIATRIVE-----LLGRRGWEALKYWNLLQYWIQEL	808
Qy	825	QKSAVSLIDTFEAVANWNTDDIILGIQIRLGRGILNIPRRVROGFERSLL	873
Db	809	KNSAVSLNLTATAVAEGTDRVLEVLQRAYRAILHPIRTIROGLERALL	857
<p>RESULT 6</p> <p>US-09-134-075-30</p> <p>Sequence 30, Application US/09134075</p> <p>Patent No. 6042836</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Berman, Phillip W.</p> <p>APPLICANT: Nakamura, Gerald R.</p> <p>TITLE OF INVENTION: HIV Envelope Polypeptides</p> <p>NUMBER OF SEQUENCES: 33</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: McCutchen, Doyle, Brown &amp; Enersen, LLP</p> <p>STREET: 3 Embarcadero Center</p> <p>CITY: San Francisco</p> <p>STATE: CA</p> <p>COUNTRY: USA</p> <p>ZIP: 94111</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Diskette</p> <p>COMPUTER: IBM Compatible</p> <p>OPERATING SYSTEM: DOS</p> <p>SOFTWARE: FastSeq for Windows Version 2.0</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/134,075</p> <p>FILING DATE:</p> <p>CLASSIFICATION:</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: 08/448,603</p> <p>FILING DATE:</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Haliday, Emily</p> <p>REGISTRATION NUMBER: 38903</p> <p>REFERENCE/DOCKET NUMBER: 14918-704</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: 415-393-2000</p> <p>TELEFAX: 415-393-2286</p> <p>TELEX:</p> <p>INFORMATION FOR SEQ ID NO: 30:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 857 amino acids</p> <p>TYPE: amino acid</p> <p>STRANDEDNESS: single</p> <p>TOPOLOGY: linear</p> <p>US-09-134-075-30</p>			
<p>Query Match 48.6%; Score 2248; DB 2; Length 857;</p> <p>Best Local Similarity 49.3%; Pred. No. 3e-176;</p> <p>Matches 438; Conservative 161; Mismatches 238; Indels 52; Gaps</p>			
Qy	5	MRAMG-KRNRK-----LGILYIVMALIIPCLSSQLYATYVAGVPWEDAAPVLCF	54
Db	1	MRVKGIRRNQHLWRWGTMLGLIMI-----CSAAGKLWTVYGVVPMKRTTTLFC	53
Qy	55	ASDANLTSTEKNVWASQACVPTDPHXYLLNVTDFNINWENYVVEQMEDIIISLWDQ	114
Db	54	ASDAKAYDIEIHNWATHACVPTDPNPQEVLENVTENFMKNVWVEQMHEDIISLWDQ	113
Qy	115	SLKPCIQMTFCIQMNCCTDKNNNTSGTENRTSSSNPMKTCFENITTVLKDKKKKQAL	174
Db	114	SLKPCVKLTPLCTNLCTDAGTNTNTSSSREKLEKEIKNCSFNITSVTRDKMQKETAL	173
Qy	175	FYVSDLTKLAD---NNTTN-TWYTLNCSNTTIKQACPKVSPFPIIYCAPAGYAFRC	230
Db	174	FNKLDIVPDDDRNSTRNSTRLISCSNTSVITQACPKVSPFPIIHFCTPAGFALLKC	233
Qy	231	NSAEFNGTGRCSISVVTCHGKIPVSTOLIINGTLKSEKIRMGKNTSDSKNIIVTL	290
Db	234	NNKTFNGSGPKNSVVOCTHGRPVVSTQILLNGSLAGEVIRSENFTNNAKTIIVQL	293
Qy	291	SSDIEITCVRPGNNQVQEMKIGP-MAWYSMALGTGNSRVAICYQNTTEWEKALKNTA	349
Db	294	TEPVKINCTRP-NNNRKSIPIGFGRAFY--ATGDIIGNIRQAHCNLSRTDW----NNTL	346
Qy	350	ERYLEINNTGNTTIFNRSQDGSVEYTHLFNFCHEGFFYCNTESEMYTF----LCN	405
Db	347	GOIVEKLREQFGNKTIFNHS-SGGDPFPIYMHFNCGEYFCTTQLEDSTWDNTKVS	405
Qy	406	GTWNCNTNSANGMTPCKLKOVRSWRRGSGSLVAPPNGLTCISHTIGMILQMDAP	465
Db	406	GTSTENSIT-----LPCKIKQIVNMQVGVKAMTAPPIRQIRCSSNITGLLITRDGG	460
Qy	466	WNKTEF-TFRPIGDMKDWRNELFKYKVRVPFSPVAPTPIARPVIGTGTREKRAVGL	524
Db	461	SNNSMNETRPGGDMNRSELYKYKVKIEPLGVAPTKARVY----QREKRAVGI	516
Qy	525	GMLFLGVLSTAGTGAATATLVQVHSVTKIGVQOQDNLRLRAIQAOQELLRLSVGIRQ	584
Db	517	GAFLFLGVLSTAGTGAATATLVQVHSVTKIGVQOQDNLRLRAIQAOQELLRLSVGIRQ	576
Qy	585	LRRLALEFLIQOQLNLNWCCKGRILCTSVYKWNENWNTNINQINGNLTWQEDQOQ	644
Db	577	LQARVLAVERYLDDQLLQGWGSGGLICTTSPVWNASN-KSLDKINDNMTWMEWRE	635
Qy	645	IDNVSSTIYEEIKAOVQOQNEKLLLEDEWASLNWLDITKWLVIKIAIIVGALIG	704
Db	636	IENYTSILYTLIESQOQNEQDLELDOWASLNWNSITAWLYIKIIFIMIVGLVG	695
Qy	705	VRIVMVLNLRVNIROGYOPLSLQIFTRQOSEAETPGRTGEGGDEGRPLIPSPQGLP	764

```

Db 174 FNKLDIVPDDDRNRSTNRSLISNTSVITQACPKVSFEPIPHCTPAGFALLKC 233
Qy 231 NSAENFGKCSNLSVVTCTHGIKFTVSTQILNGTSLKIRIMGNISDSKNIIVTL 290
Db 234 NKKFNGSGCKNVSTVOCTHGIKFTVSTQILNGTSLKIRIMGNISDSKNIIVTL 293
Qy 291 SSDIEITCVRPNGNQTVQEMKIGP-MAWYSMALGTGNSRNRVAYCOVNTWEKALKNTA 349
Db 294 TEPVKINCRP-NNTRKSPIGPGRAF--ATGDIIGNRQAHCHNLSRDM-----NNTL 346
Qy 350 ERYLELNNTEGNTMIFNPSQDSQDVVEVTHLHFNCHGEFFYCNTESEMYTF-----LCN 405
Db 347 GQIVKLEUREFGNKTIIFNHS-SGGDEPIVHFNCHGEFFYCNTESEMYTF-----LCN 405
Qy 406 GTNCNNTQSNISANGMIPCKLKOVVRSMRGGSLVAPPIPGNLCTISHTGMILQMDAP 465
Db 406 GTSTEENSTIT-----LPCRIKQIVNMWQEVGKAMVAPPIRGQIRCSNITGLLTRDGG 460
Qy 466 WNKTEN-TFRPIGDMKDINRNELFKYVYVVPFVSAPTPIARPVIGTGTREKRAVGL 524
Db 461 SNNSMNETFRPGGDMRDNRSERYKVKVIEPLGVAPTAKRRV-----QREKRAVGI 516
Qy 525 GMLFGLVLSAAGSTMGAATATVQHSVKGIVQOQDNLRAIQAOQELLRLSWGIRQ 584
Db 517 GAVFLGFLGAGSTMGAASTLTQVQARLLSGIVQOQDNLRAIQAOQELLRLSWGIRQ 576
Qy 585 LRARLALETIONQOLLMLWCKGLICYSVKNWETWRNTNINQWGLNLTWQENDQ 644
Db 577 LQARVLAVERYLQDQLLWCGSGKLICTSVPNWASHN-KSLDKIWDNNTWMEWERE 635
Qy 645 IDNVSITIEETOKAQVOEQNEKLELDEWASLWNNLDTIKWYIKIAIIVGALIG 704
Db 636 IENVTSLITLESQOQNEKLELDEWASLWNNLDTIKWYIKIAIIVGALIG 695
Qy 705 VRIVMVLNVRNIRGQOPLSLQIPTROQSEAPTPGRTGEGGDEGRPLIPQGFPLP 764
Db 696 LRIVFAVLIVNRGQISPLSFQRLPAPRPRDPEIEEGEGGQDRSIRLVQDGLA 755
Qy 765 LLYTDLRTILWYHLLSNLSGTQTVISHURLGLWILGKIIDACRICAAVHWLQEL 824
Db 756 LWDLLRSCLFSYHRLDILLIATRIYE-----LLGRGWELKYNWLLQYWLQEL 808
Qy 825 QKSATSLDITFAVANNTDDIILGIGRILGRILNIPRVQGFERSIL 873
Db 809 KNSAVSLNNTAIAVAGTDRVLEVLQRAYRAILHPIRIRQGLERALL 857

```

RESULT 7

US-08-448-603A-28

Sequence 28, Application US/08448603A

Patent No. 5864027

GENERAL INFORMATION:

APPLICANT: Berman, Phillip W.

APPLICANT: Nakamura, Gerald R.

TITLE OF INVENTION: HIV Envelope Polypeptides

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP

STREET: 3 Embarcadero Center

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

\* MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/448,603A

FILING DATE: 07-JUN-1994

CLASSIFICATION: 435

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/072,833
; FILING DATE: 07-JUN-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Halliday, Emily
; REGISTRATION NUMBER: 38903
; REFERENCE/DOCKET NUMBER: 14918-704
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-393-2000
; TELEFAX: 415-393-2286
; TELEX:
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-448-603A-28

```

Query Match 48.6%; Score 2247; DB 2: Length 850;  
Best Local Similarity 50.2%; Pred. No. 3.5e-176;  
Matches 432; Conservative 157; Mismatches 236; Indels 36; Gaps 14;

```

Qy 20 IVALLIPCLSSQLYATVYGVVWEDAAPVLFASDANLSTSEKHNYWASQACVPTDP 79
Db 19 MLLGLMICAESAELKVTYGVVWKEATTLFASDADAKAYDTEVHNWATHACVPTDP 78
Qy 80 TPHEYLITNTVDFNFWYEQEDIIISLWDSQSLKPCIOFTWFCIOMNCTDIKN--N 137
Db 79 NPQIEIGLVNTFNWKNWYEQMHEDIIISLWDSQSLKPCVLTPLCVLTNCTDKNATN 138
Qy 138 NTSGTENRTSSSENPMKTCFNTITVLKDKKKEKQALFYVSDLTKLADNNTNTWTNLTN 197
Db 139 TTSSSGMKMERGE--IKNSFNWTSIRDKMNEVALFYKLDVVP1-DND--NTSYRLIS 193
Qy 198 CNSTTIKQACPVSEPIIYACAGYAIKFCNSAEFNGTKCSNISVVTCTHGIKFTV 257
Db 194 CNTSVITQACPVSEPIIYACAGYAIKFCNSAEFNGTKCSNISVVTCTHGIKFTV 253
Qy 258 STQILLNGTSLSEKIRIMGNISDSKNIIVTSLSDIEITCVRPNGNNTVQEMKIGP-MA 316
Db 254 STQILLNGSLAEVEEVIRSAFSDNAKTIIVQLNESVEINCRP--NNTRSIHIGPRA 312
Qy 317 WYSMALGTGNSRNRVAYCOVNTWEKALKNTAERYLELNNTEGNTMIFNRSQDSDV 376
Db 313 FY--ATGEIIGDIRQAHCHLSSTKNWNTLKQIVTKLREHFNK-----TIVFNHS-SGSDP 364
Qy 377 EYTHLHFNCHGEFFYCNTESEMF-----NYTFLCNGTNCNNTQINSANGMIPCKLKQVVR 432
Db 365 EIVMHSFNGCGEFFYCNTESEMF-----NYTFLCNGTNCNNTQINSANGMIPCKLKQVVR 421
Qy 433 WMRGSGLYAPPIPGNLCTISHTGMILQMDAPWNKTENTFRPIGGDMKDIWRNLFKYK 492
Db 422 WQEVGKAMTAPPIRGQIRCSNITGLLTRDGGNNSETEIFRPGGDMRDNWSELYKYK 481
Qy 493 VVRVPKPSVAPTPIARPVIGTGTREKRAVCLGMLFGLVLSAAGSTGAATAALTQVTHS 552
Db 482 VVKEPLGVAPTAKRRV-----QREKRAVIGAVFGLGAGSTGAATAALTQVTHS 537
Qy 553 VIKGIVQOQDNLRAIQAOQELLRLSWGIRLRAILLALLETIQOQLNLWCGKRLI 612
Db 538 LLSGIVQOQDNLRAIEAQHLLQLTWGIKQLQARVLAVERYLKQQLLGLWCGSGKLI 597
Qy 613 CYTSVKWNETWRNTNINQWGLNLTWQENDQDNDVSSYIIEEIQAKVQOQEQNEKLE 672
Db 598 CTAVPNASHN-KSLDKIWDNNTWMEWEREIDNTSLIYSLIEESQOQNEKLE 656
Qy 673 LDWASLWNNLDTIKWYIKIAIIVGALIGVIRVIMVLNVRNIRGQVQPLIQIPTR 732
Db 657 LDKWASLWNNLDTIKWYIKIFIMVGLVGLVIFVTVLSIVNRVKGYSPLSFOTHP 716
Qy 733 QQSEAEPTGRTGEGGDEGRPLIPSPQGFPLPLYTDLRTIILWYHLLSNLSISGTQTVI 792

```



Db 717 APRGLDRPEGEGERDRSRRLVDGFLAIYVWDLRSCLFSYHRLDRLIAARIY 776  
QY 793 SHRLGLWILGOKIIDACRICAAVHYWLOKQSAATSLDTEFAVANWTDLILGIOR 852  
Db 777 E-----LGRGWEALKYWNLLQYWIQELKNSAVSLNATAIAVAEGTDRVIEIVQR 829  
QY 853 LGRGILNIPRRVROGFERSLL 873  
Db 830 AYRAILHIPTRIOGLERALL 850

RESULT 8  
US-09-134-075-28  
; Sequence 28, Application US/09134075  
; Patent No. 6042836  
; GENERAL INFORMATION:  
; APPLICANT: Berman, Phillip W.  
; APPLICANT: Nakamura, Gerald R.  
; TITLE OF INVENTION: HIV Envelope Polypeptides  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McCutchen, Doyle, Brown & Ebersen, LLP  
; STREET: 3 Embarcadero Center  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/134,075  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/448,603  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haliday, Emily  
; REGISTRATION NUMBER: 38903  
; REFERENCE/DOCKET NUMBER: 14918-704  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-393-2000  
; TELEFAX: 415-393-2286  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 850 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-134-075-28

Query Match 48.68; Score 2247; DB 3; Length 850;  
Best Local Similarity 50.28; Pred. No. 3.5e-176;  
Matches 432; Conservativity 157; Mismatches 236; Indels 36; Gaps 14;

QY 20 IVMAIIPCLSSQLYATYVAGVPWEDAAPVLFCSADANLTSTKENVWASQACVPTDP 79  
Db 19 MLLGMLMCSAAELWTVYGVFWKEATTLFCASDAKAYDEVHNVWATHACVPTDP 78  
QY 80 TPHEYLNTVNTNINVENYVMEQMEDIIISLWQSLKPCIQMTFMCIQMCTDIKN--N 137  
Db 79 NPQSIGLENVTENFMKNMNNVMEQMEDIIISLWQSLKPCVKLPVLCVTLNCTDLKNATN 138  
QY 138 NTSGTENRTSSSNPMKTCENITVILKDKKQALFYVSDILKLDANNNTNTMYTLIN 197  
Db 139 TTSSWGKMERGE--IKNCSENVTTISDKMNEYALFYKLDVVYPI-DND--NTSYRLIS 193

QY 198 CNSTTIQACPKYSEFPPIYIYCAPAGYALFKONSAEFGNGTKSCNISVYVTCFHGKPTV 257  
Db 194 CNTSVITQACPKYSEFPPIYIYCAPAGYALFKONSAEFGNGTKSCNISVYVTCFHGKPTV 253  
QY 258 STOLLINGTSLKIRIMGNISDSGNIIIVTILSSDIEITCVPRGNQTVQEMKIGP-MA 316  
Db 254 STOLLINGSAAEEVIRSANFSDNAKTIIVQLNESVEINCTRP--NNNTRRSIHGPGRA 312  
QY 317 WYSMALGTGSRNRVAYCOYNTTEWEKALKATAERYLELINNTEGNTMTLNFNSQDSDV 376  
Db 313 FY--ATGEIIGDIOAHCNLSSTKWNNTLQIVTKLREHFNK-----TIVFNIS--SGDP 364  
QY 377 EVTHLHFNCHGEFFYCNTESEMP-----NYTFLCNGTNCNNTQNSANGMIPCKLKQVRS 432  
Db 365 EIVMHGFNCGGEFFYCNTESEMP-----NYTFLCNGTNCNNTQNSANGMIPCKLKQVRS 421  
QY 433 WMRGSSGLYAPPINLTCLSHITGMLQMDAPWNTKENTFRPIGGDKMDIWNELPKYK 492  
Db 422 WQEVGKAMYAPPINLTCLSHITGMLQMDAPWNTKENTFRPIGGDKMDIWNELPKYK 481  
QY 493 VVRKPSVAPTPARPIVIGTGHREKRAYGLGMLFLGVLSSAAGSTMGAAATALTVOHIS 552  
Db 482 VKIEPLGVAPTAKRRYM-----QREKRAYGIGAVFLGLGAAGSTMGAAATALTVOHIS 537  
QY 553 VIRGIQQQDNLRAIQAOQELRLSVGIRQLRALETLIQOQLLNLMLGCKGRLI 612  
Db 538 LLSGIVQQNNLLRAIEAQHLLQLTWGIKQLQARVLAVERYLKQDQLLGIWGCCKLI 597  
QY 613 CYTSVKNWETWNTNINQWNLTWQEWQDQIDNVSSSTIYIEIQRAVOQOENKELLE 672  
Db 598 CTTAVPWNASWN--KSLDKIWNNTWNEIREIDNTSLIYSLEESQNOQKNEQELLE 656  
QY 673 LDEWASLWNLDTIKWLYIKIIVGALIGVIRVIMVLNVRNIROGYOPLSLQIPTR 732  
Db 657 LDKWASLWNLDTIKWLYIKIIVGALIGVIRVIMVLNVRNIROGYOPLSLQIPTR 716  
QY 733 QQSEAEPTGRTGEGGDEGRPLIPSPQGFPLLYTDLRTIILWSYHLLSNLSGTQTVI 792  
Db 717 APRGLDRPEGEGERDRSRRLVDGFLAIYVWDLRSCLFSYHRLDRLIAARIY 776  
QY 793 SHRLGLWILGOKIIDACRICAAVHYWLOKQSAATSLDTEFAVANWTDLILGIOR 852  
Db 777 E-----LGRGWEALKYWNLLQYWIQELKNSAVSLNATAIAVAEGTDRVIEIVQR 829  
QY 853 LGRGILNIPRRVROGFERSLL 873  
Db 830 AYRAILHIPTRIOGLERALL 850

RESULT 9  
US-08-022-835-2  
; Sequence 2, Application US/08022835  
; Patent No. 5420030  
; GENERAL INFORMATION:  
; APPLICANT: Reitz Jr., Marvin S.  
; APPLICANT: Franchini, Genoveffa  
; APPLICANT: Markham, Phillip D.  
; APPLICANT: Gallo, Robert C.  
; APPLICANT: Lori, Franco C.  
; APPLICANT: Popovic, Mikulas  
; APPLICANT: Gartner, Suzanne  
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: Eleventh Floor, 1615 L. Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

QY 5 MRAMG-KRNRK--LGILYIVMALIIPCUSSQLYATVYAGVPVWEDAAPVLFCSANDLNT 61

Query Match	47.8%	Score 2211.5	DB 1	Length 856
Best Local Similarity	49.4%	Pred. No. 2.9e-173		
Matches 440	Conservative 159	Mismatches 236	Indels 55	Gaps
QY	5	MRAMG-KNRK--LGILIVMALIIPCLSSQLYATYAGVPVWEDAAPVLFASDANLT	61	
DB	1	MRVKIRNYYOHWWGTMFLJGLMIGSATEKLMVTIVYGVVYVKEATTLFLCASDAKAY	60	
QY	62	STEKNHWASOACVPTDTPHEYLLTNTVDNFNLWENKYNVEQMEDIIISLDOSLPCQI	121	
DB	61	DEYHNWATOCVPTDPENQOEVLVNTNFNWKNNKMYEQHEDIIISLDOSLPCVK	120	
QY	122	MTFPCIQMNCCTDIKN--NTSGTENRTSSFE----NPMKTCFNTITVLKDKKKQAL	174	
DB	121	LTPCLVTJLCTDLRNTTNTNNSNNEGTLKGMKNCFSNITTSIRDMQKEYAL	180	
QY	175	FVSDLTKLADNNTNTMYTLINCSTTIQACPKVSFPEPIIYCAPAGYAIKCNSE	234	
DB	181	LYKLDIVSI-DNOSTS--YKLISNTSVITQACPKISFPEPIIYCAPAGFAILKCNCK	237	
QY	235	FNGRGKCSISVVTCTHGIRKPTVSTOLLNGLTSLEKIRIMGNKNTSDSKNIIIVLSSDI	294	
DB	238	FSRGSKCNYSTVOCTHGIRPVYSTQLLNGSLAEVEVIRSENTDNNKTIIVHLSNV	297	
QY	295	EITCVREGNQTVQEMKIGP-MAWYSN--ALGTGSNRSRVAYCOYNTTEWEKALKANTAER	351	
DB	298	QINCTRPYNNKR-KRTHIGPGRAYFTTKNIIGT---IROAHCNISRAKWDTLRQIYSK	352	
QY	352	YLELINNTEGNTMIFNRSDGSDVBVTHLHFNCHGEFFYCNTESEMPYTFLCNGTNC-N	410	
DB	353	LKEQFK---NKTIVFNQS-SGDDPEIYVMSFNCGEFFYCNTESEMPYTFLCNGTNC-N	405	
QY	411	NTOSINSANGMTPCKLKQVVRSMRGSGSLYAPPIPGNLTICSHITGMILQ----MDAPW	466	
DB	406	NTGSSNN-NITLQCKIKQIINHWQEVGKAMYPPIEGQIRCSSNITGLLLTRDGGKDTDT	464	
QY	467	NKTENTFRPIGGDMKDWRNELFKYVVRVPFVSAPYPIARPVIGTGTREKRAVGLGM	526	
DB	465	NDTE-IFRPGGDMRDNRSLEYLYKVVYTIPELGVAPYAKRRVY----QREKA-AIGA	518	
QY	527	LFGLVLSAGSTWGAATAFTVOTHSVIKGVQOQDNLLRAIOAQQLRLRSVWGIRQLR	586	
DB	519	LFGLFLGAAGSTWGAASVTLTQARLLLSGVQQQNLLRAIEAQQHWMLQITVWGIKQLQ	578	
QY	587	ARLLAETLIONQOQLLNLCWCKGRLCYTSVKWNTEWRTNTNINQINWNTJWQFWDQIID	646	

Db 1 MRVGIIRNYOHWWGWTMLGLLMICSADEKLVTVYGVVWKEATITLFCASDAKAY 60  
Qy 62 STEKHNVWASQACVPTDPTPEYLLTNVDFNIWENYVWQEDIIISLDQSLKPCIQ 121  
Db 61 DTEVHNWATQACVPTDPNPEVELYVNFNWKNNVWQEHDIISLDQSLKPCVK 120  
Qy 122 MTFWCIOMNCTDIKN--NNTSGTENRTSSSE-----NPMKTCFNITTVLKDKKKEKQAL 174  
Db 121 LTPCLVTLNCTDURNTNTNNSANNSEGTIKGEMKNCNFNTTIRDRKQKEYAL 180  
Qy 175 FVYSDTLKADNNTMTYTLNCTNTTKACPKVSEFPIYIYCAPAGYAFKCNSE 234  
Db 181 LYKLDIVSI--DNST--YELISCNVITQACPKISFEPIYIYCAPAGYAFKCNKDK 237  
Qy 235 FNGTGKSNISVYCTHGIKPTVSTQILNGLTSLKIRIMGNKNSDGNKIIVTSSDI 294  
Db 238 FSGGCKNVSTVQCTHGIKPTVSTQILNGLTSLKIRIMGNKNSDGNKIIVTSSDI 297  
Qy 295 EITCVRPNNTVQEMKIGP--MAYISM--ALGTGNSRSRVAYCOYNTTEWELKNTAER 351  
Db 298 QINCTRPYNKR--KRIHIGPGRAFYTKNIIGT----IROAHCNISRAKWNWTLRQIVSK 352  
Qy 352 YLELINTEGNTMTIFNRSQDSDVEVTHLHFNCHEFFCYNTSEMFNTFLCNGTNC-N 410  
Db 353 LKEQFK----NKTIVFNQS--SGGDPVIMHFNCGGFEFFCYNTSPLFNSTW--NGNNTW 405  
Qy 411 NTSINSANGMPCPKLQVVRWNRGSGLYAPPPIGNLTCISHTGMILQ----MDAPW 466  
Db 406 NTTGSNN--NITLOCKIKQIINNWEVEGKAMYAPPIEGQIRCSNITGLLITRDGKDTDT 464  
Qy 467 NKTENFRPIGDKMDIWRNELFKYVVRKVPFSAVPTIARPIVIGTHREKRAVLGM 526  
Db 465 NDTE--IFRPGGDMRDNWRSSELYKVKVYIEPLGVATPKAKRRVY---QREKRA-AIGA 518  
Qy 527 LFLGVLNAGTGAATATLVQVSHVKGIVQOQDNLRAIQAOQELLSLVWIGRQLR 586  
Db 519 LFLGFLAAGSTGAASVTLTVQARLLSGIVQOQDNLRAIQAOQELLSLVWIGRQLR 578  
Qy 587 ARLLALETIONQOLLNMLGCKRLICVTSVKWNETWRNTNINOINWNTWQEDD 646  
Db 579 ARVLAVERLKDQQLLGFMGSGKGLICTITVPVWASNS--KSLDDIWNNTWQWEREID 637  
Qy 647 NVSSTIIEIOKAQVQEQNEKLELDEWASLWNLDDITKWLVIKIIIVGALIGVR 706  
Db 638 NYSLSLLEKSTQEQNEKLELDEWASLWNLDDITKWLVIKIIIVGALIGVR 697  
Qy 707 IVMIVLNRNIGROYPLSLQIPTRQOSEAETPGRTGEGGDEGRP---RLIPSPQGL 763  
Db 698 IVFVAVLSVNRVROGYSPLSLQIPTRPVPGRDPEGIEEGGERDRDTSGRLV---HGFL 754  
Qy 764 PLLYTDLRTILWYSHLLSNLSGTQVISHRLGLWLGKIIDACRICAAVHYWLOE 823  
Db 755 AIIWVDLRSFLSYHRDLILLIARIV-----ELGRRGWELVKYWNLLQYWSQE 806  
Qy 824 LQKATSILDTFAVANWTDIILGIQRLGRGILNTPRRVQCFERSLL 873  
Db 807 LKSSAVLLNATAVAEGTDRVLEVLQAGRAILHIFTRIQGLERAIL 856

## RESULT 11

US-08-647-714-2

Sequence 2, Application US/08647714

Patent No. 5869313

GENERAL INFORMATION:

APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,

APPLICANT: GENOVEFA; MARKHAM, PHILIP D. GALLO, ROBERT

APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND

APPLICANT: GARTNER, SUZANNE

TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK, 3.5"  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/647,714  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/388,809  
FILING DATE: 15-FEB-1995  
APPLICATION NUMBER: US 08/022,835  
FILING DATE: 23-FEB-1993  
APPLICATION NUMBER: US 07/599,491  
FILING DATE: 17-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: LESLIE A. SERUNIAN  
REGISTRATION NUMBER: 35,353  
REFERENCE/DOCKET NUMBER: 2026-4092US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 856 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-647-714-2

Query Match 47.8%; Score 2211.5; DB 2; Length 856;

Best Local Similarity 49.4%; Pred. No. 2.9e-173;

Matches 440; Conservative 159; Mismatches 236; Indels 55; Gaps 23;

Qy 5 MRAMG-KRNRK--LGILYIVMALIIFCLSSQLYAVIYAGVPVWEDAAAPVLCASDANLT 61  
Db 1 MRVGIIRNYOHWWGWTMLGLLMICSADEKLVTVYGVVWKEATITLFCASDAKAY 60  
Qy 62 STEKHNVWASQACVPTDPTPEYLLTNVDFNIWENYVWQEDIIISLDQSLKPCIQ 121  
Db 61 DTEVHNWATQACVPTDPNPEVELYVNFNWKNNVWQEHDIISLDQSLKPCVK 120  
Qy 122 MTFWCIOMNCTDIKN--NNTSGTENRTSSSE-----NPMKTCFNITTVLKDKKKEKQAL 174  
Db 121 LTPCLVTLNCTDURNTNTNNSANNSEGTIKGEMKNCNFNTTIRDRKQKEYAL 180  
Qy 175 FVYSDTLKADNNTMTYTLNCTNTTKACPKVSEFPIYIYCAPAGYAFKCNSE 234  
Db 181 LYKLDIVSI--DNST--YELISCNVITQACPKISFEPIYIYCAPAGYAFKCNKDK 237  
Qy 235 FNGTGKSNISVYCTHGIKPTVSTQILNGLTSLKIRIMGNKNSDGNKIIVTSSDI 294  
Db 238 FSGGCKNVSTVQCTHGIKPTVSTQILNGLTSLKIRIMGNKNSDGNKIIVTSSDI 297  
Qy 295 EITCVRPNNTVQEMKIGP--MAYISM--ALGTGNSRSRVAYCOYNTTEWELKNTAER 351  
Db 298 QINCTRPYNKR--KRIHIGPGRAFYTKNIIGT----IROAHCNISRAKWNWTLRQIVSK 352  
Qy 352 YLELINTEGNTMTIFNRSQDSDVEVTHLHFNCHEFFCYNTSEMFNTFLCNGTNC-N 410  
Db 353 LKEQFK----NKTIVFNQS--SGGDPVIMHFNCGGFEFFCYNTSPLFNSTW--NGNNTW 405  
Qy 411 NTSINSANGMPCPKLQVVRWNRGSGLYAPPPIGNLTCISHTGMILQ----MDAPW 466  
Db 406 NTTGSNN--NITLOCKIKQIINNWEVEGKAMYAPPIEGQIRCSNITGLLITRDGKDTDT 464

```

QY 467 NKTENTFRPIGGDMKDIWRNELEFKYKVVVVPFVAPTPVIGTCTHREKRAVGLGM 526
Db 465 NDTE-IFRPGGDMRDNWSELYKVKVTIEPLGVAPTAKRRVV----QREKRA-AIGA 518
QY 527 LFGVLSAAGSTMGAAATALTQTHSVIKGIVQOQDNLRLAIOAQOELLRLSVGIRQLR 586
Db 519 LFLGFLGAAGSTMGAASTVLTQVQALLSGVQOQNNLLRAIEAQOELMLQLTVMGIRQLQ 578
QY 587 ARLLALETIONQOOLLNMGCKGRILCYTSVKWNETWNTTINQINWNTWQOQOIID 646
Db 579 ARVLAVERLYKQOQLLFGWGSGLKCTTTPVWNASNS-KSLDDIWNNTWQWEREID 637
QY 647 NVSTYIEIQKAOVQOQBNKLELDEWASLNNWLDITKWLVIKIAIIVGALIGVR 706
Db 638 NYSLSIYSLKESQOQKNEQELLELDKWSLNNWDTNWLWIKIFIMVIGGLVGLR 697
QY 707 IVMIVLNVIRIOGYQPLSQITPQOQSEATPGRTEGGGDEGRP----RLIPSPQGL 763
Db 698 IVFVLSIVNRVROGYSPLSQITPVPVPGDRPEGTEEGGERDRDTSGLV----HGFL 754
QY 764 PLLYTDLRTILWYHLLSNLSIGTQVISHRLGLWILGOKIIDACRCAAVIHWLQ 823
Db 755 AIIWDLRSLFLSHRDLLIARIV------ELGRRGWEVLYKWNLLQWQOE 806
QY 824 LOKSATSILDTFAVANWTDITLILGRLGRGILNIPRVQGFERSLL 873
Db 807 LKSAVSLNATAVAAGTDRVIEVQLORAGRAILHIPRIQGLERALL 856

```

RESULT 12

```

us-08-463-210-11
; Sequence 11, Application US/08463210
; Patent No. 6001977
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAI, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; * COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,210
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 436
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6949
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 863 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..863
; OTHER INFORMATION: /note= "env protein of HTLV-III"
;
us-08-463-210-11

```

```

Query Match 47.8%; Score 2211.5; DB 3; Length 863;
Best Local Similarity 49.3%; Pred. No. 3e-173;
Matches 427; Conservative 164; Mismatches 238; Indels 33; Gaps 16;
QY 20 IVMALIIICLSSQLXATVYAGVPVWEDAAPVLFCAASDANLSTKHNWYSAQACVPDP 79
Db 27 MLLGLMICSATEKLVVYVYGVVWKEATTLFCAADAKAYDEVHNVNATHACVPTDP 86
QY 80 TPHEYLITNTDNTFNWYVMEQOEDIISLWDSLAPCIQMTFCIQMNCITKNN-N 138
Db 87 NPQEVVLNVTFENFMKNDVMEQMHEDIISLWDSLAPCVKLTPLCVSLACTLKN 146
QY 139 TSGTENRTSSSNPMKTCENITTVLKDKKQKQALFYVSDTLKADNNTNTMYTLNC 198
Db 147 TNSSSGRMIMEKEIKNCFSNISTIRKVQKEYAFFYKLDIPI-DNDTTS--YTLTSC 203
QY 199 NSTTIQACPKVSPPIPIYCAPAGYAFKCSNAEFNGTCKSNISVVTCTHGIKPTVS 258
Db 204 NTSVITQACPKVSFPIPIHYCAPAGFAILCKNNKTENGTCINVTQCTHGIKPTVS 263
QY 259 TOLLNGTLSEKIRIMKNIIVTLSSDIETCVPCNNQTVQEMKI--GPMA 316
Db 264 TOLLNGSLAEVIRSANETDNKIIIVOLNOSVEINCTRP--NNTRKSIIRIQGPR 322
QY 317 WYSMALGTGSRNRSVAYCOYNTWEKALKNTAERYLELINTEGNTMIFNRSQGSVD 376
Db 323 AF-VTIGKIGN-MQAHGHNISRAKWNNTLQIDSKLREQFN--NKTIIFKSSSGDPE 377
QY 377 EVTHLFNCHGEFFCYNTSEMFNTFELNGT----NCNWTQINSANGMIPCKLKQVRS 432
Db 378 IVTH-SFNGCGEFCYNTOLFSTWF-NSTWTKGNSNTEGSDTIT--LPCRIKOIINN 433
QY 433 WMRGSGLYAPPPIGNLTCTSHITGMILQDAPWNKTEN-TFRPIGDMKDWRNELFKY 491
Db 434 WQEVGKAYAPPISGQIRCSNITGLLLTRDGNSSNESEIFRPGGDMRDNWRSLEYKY 493
QY 492 KVRVKPFSVAPTPIARPVIGTCTHREKRAVGLMLFLGVLSAAGSTMGAAATALTQ 551
Db 494 KVKIEPLGVAPTAKRRVV----QREKRAVIGALFLGLGAAGSTMGAAATALTQ 549
QY 552 SVKGVQOQDNLRLAIOAQOELLRLSVGIRQLARLLALETLIONQOLLNMGCKGR 611
Db 550 QLLSGIVQOQNNLLRAIEAQOHLQLQVIGIKQLQARILAVERYLKDQOLGICGCK 609
QY 612 ICYTSVKWNETWNTTINQINWNTWQOQOQOQOQOQOQOQOQOQOQOQOQOQO 671
Db 610 ICTAVPWNASNS-KSLEQIWNNTWQWEDREINNYTSLIHSLEESQOQOQOQO 668
QY 672 ELDEWASLNNWLDITKWLVIKIAIIVGALIGVRIMVNLVNRIRQGYQPLSLQIPT 731
Db 669 ELDKWASLNNWNTNWLWYIKLFIMVIGGLVGLRIVFAVLSVNRVRRQGSPLSF 728
QY 732 RQOSEATEPRTGEGGDEGRPLIPSPQGLPLDLYTLDRITILWYHLLSNLSIGT 791
Db 729 PTPRGDPREGTEEGGERDRDRSLVNGSLALIDDLRLSLCFLSYHRLDILLVTRI 788
QY 792 ISHLRLGLWILGQKIIDACRCAAVIHWLQELOKSATSILDTFAVANWTDIIIGIQ 851
Db 789 VE-----LLGRGWEALKYWNLLQWSELKNSAVSLNATAVAAGTDRVIEVQ 841

```

QY 852 RLGRGILNIPRRVQGFERSLL 873  
 Db 842 GAYRAIRHPRIRQGLERILL 863

## RESULT 13

US-08-022-835-4  
 ; Sequence 4, Application US/08022835  
 ; Patent No. 5420030  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reitz Jr., Marvin S.  
 ; APPLICANT: Franchini, Genoveffa  
 ; APPLICANT: Markham, Phillip D.  
 ; APPLICANT: Gallo, Robert C.  
 ; APPLICANT: Lori, Franco C.  
 ; APPLICANT: Popovic, Mikulas  
 ; APPLICANT: Gartner, Suzanne  
 ; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1 AND USES  
 ; TITLE OF INVENTION: THEREOF  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
 ; STREET: Eleventh Floor, 1615 L. Street, N.W.  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20036-5601  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/022,835  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/599,491  
 ; FILING DATE: 17-OCT-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Scott, Watson T.  
 ; REGISTRATION NUMBER: 26,581  
 ; REFERENCE/DOCKET NUMBER: WTS/5683/78507/SRL  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 861-3000  
 ; TELEFAX: (202) 822-0944  
 ; TELEX: 6714627 CUSH  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 857 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-022-835-4

Query Match 47.8%; Score 2211; DB 1; Length 857;  
 Best Local Similarity 49.1%; Pred. No. 3.2e-173;  
 Matches 437; Conservative 159; Mismatches 240; Indels 54; Gaps 22;  
 QY 5 MRANG-RNRK--LGILYIVNALIPCLSSSOLYATYAGVPVWEDAAVLFCAADANLT 61  
 Db 1 MRVKGIRNTHQHWGWTMLGLLMICATEKLVVTVYGVVYVKEATTLFCASDAKAY 60  
 QY 62 STEKHNVASOACVPTDPTTHEYLLTNVFNWENYVMEQMEDIIISLDQSLKPCIQ 121  
 Db 61 DTEYHNVWATHACVPTDPNFQEVLEVNVTFNFWKNNVYEQMHEDIISLDQSLKPCVK 120  
 QY 122 MTFMCIONCNDIKN--NNTSGTENRTSSSE-----NPMKTCENITTVLKDKKKEQAL 174  
 Db 121 LTPCLVCLNCTDLRNTNTNNSSTANNNSSEGTIKGEMKNCNNTITSIRDKMQKREYAL 180  
 QY 175 FYVSDLTKLADNNTNTMTYLINCNSTTIKQACPVKSFEPPIYCAPAGYAFKCSAE 234

Db 181 LYKLDIVSI---NDSTSYRLISCNSTVITQACPKISFEPPIHYCAPAGFAILKCNDDK 237  
 QY 235 FNGTGKCSNISVYCTHGIKPTVSTOLILNGTSLSKKIRIMKGNISDSGKNIIVTLSSDI 294  
 Db 238 FSGKSGCKNVSTVOCTHGIKPTVSTOLILNGTSLSKKIRIMKGNISDSGKNIIVTLSSDI 297  
 QY 295 EITCVRPNNQTVQEMKIGP--MAWISM--ALGTGSRNRSRVAYQYQYNTTEWEKALKNTAER 351  
 Db 298 QINCTREYNKR-KRIHIGPGRAFYTKNIIGT---IROAHCNISRAKWNIDILQIVSK 352  
 QY 352 YLELINNTEGNTMIFNRSQDSDVEVTHLHFNCHGEFFCYNTSEMPTYFLNGTNC-N 410  
 Db 353 LKEQFK---NKTIVFNQS--SGDPEIVMHSFNGCGEFCYNTSPLFNSTW--NGNNTWN 405  
 QY 411 NTOSINSANGMIPCKLKQVYRSMRGSGLYAPPICNLNLCISHTGMIQ---MDAPW 466  
 Db 406 NTGSSNN-NITLQKIKQIINMMQEVGKAIYAPPIEQIRKCSNITGLLLTRGGKDTDT 464  
 QY 467 NKTENTFRPIGDMKDIWRNELEFKYVYRVKPSVAPTPIARVIGTGTREKRAVGLGM 526  
 Db 465 NDTF-IFRPGGDMRDNRSELYKYVVTIEPLGVAPTAKRRV---QREKRA-AIGA 518  
 QY 527 LFLGVLSAGSTMGAAATALTVOHSHVIGIVQOQNLRLRAIQAOQLLRLSVGIRQLR 586  
 Db 519 LFLGFLGAGSTMGAAATALTVOHSHVIGIVQOQNLRLRAIQAOQLLRLSVGIRQLR 578  
 QY 587 ARILALETIONOGLNLWGCKGRLCYTSVKWNETWRNTNINOIWNGLTWOEWDOQID 646  
 Db 579 ARILAVERYLKDQQLLIGWCSGLKICTITVPWNASWN-KSLDDIWNNTWOWEREID 637  
 QY 647 NVSSTIYEIQKAVQOQNEKKLEDEWASLWNLMDITKWLWIKIAIIVGALIGVR 706  
 Db 638 NYSLSIYLSLEKSTQOQENNEQELLELDKWSLWNLMDITKWLWIKIAIIVGALIGVR 697  
 QY 707 IMVNLVNRNIRQGYOPLSLQIPTRQOQSEATPGRTGEGGDEGRP---RLIPSPQGL 763  
 Db 698 IYFAVLISVNRVQGYSPLSLQIPTRQOQSEATPGRTGEGGDEGRP---HGFL 754  
 QY 764 PLAYTDLRTIILWSYHLNLSIGTOTVISHLRLGLWILGOKIIDACRICAADVHYWLOE 823  
 Db 755 AILWDLRLSLFSLYHLRLDLLIAARIVE-----LLGRGWEVLKYWNLLQYWSOE 807  
 QY 824 LOKSATSLIDTAVAVANWTDIILGILQIRGLRGILNIPRRVQGFERSLL 873  
 Db 808 LKSSAVSLNATDIAVAGTDRVIEVQLQAGRAILHPTIRQGLERALL 857

## RESULT 14

US-08-388-809-4  
 ; Sequence 4, Application US/08388809  
 ; Patent No. 5576000  
 ; GENERAL INFORMATION:  
 ; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,  
 ; APPLICANT: GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT  
 ; APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND  
 ; APPLICANT: GARTNER, SUZANNE  
 ; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1  
 ; TITLE OF INVENTION: AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MOHGAN & FINNEGAN  
 ; STREET: 345 PARK AVENUE  
 ; CITY: NEW YORK  
 ; STATE: NEW YORK  
 ; COUNTRY: USA  
 ; ZIP: 10154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: FLOPPY DISK, 3.5"  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/388,809  
 FILING DATE: 15-FEB-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/022,835  
 FILING DATE: 25-FEB-1993  
 APPLICATION NUMBER: US 07/599,491  
 FILING DATE: 12-OCT-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: LESLIE A. SERUNIAN  
 REGISTRATION NUMBER: 35,353  
 REFERENCE/DOCKET NUMBER: 2026-4092US2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 751-6840  
 TELEFAX: (212) 751-6849  
 TELEX: 421792  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 857 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-388-809-4

Query Match 47.8%; Score 2211; DB 1; Length 857;  
 Best Local Similarity 49.1%; Pred. No. 3.2e-173;  
 Matches 437; Conservative 159; Mismatches 240; Indels 54; Gaps 22;

QY 5 MRAMG-KRNRK--LGLIYVMAIIPCLSSSQLYATYAGVPVWEDAAPVLFCSANLIT 61  
 DB 1 MRVKGIRRYQHGWGTLGLMISATEKLMVTVYGVVPVKEATITLFCASDAKAY 60

QY 62 STEKHNVASQCVPTDTPHEYLITNTVNFNIWENYEQMEDIIISLWQSLKPCIQ 121  
 DB 61 DTEVHNVATHACTPTDPNPOEVLNVTFENNMKNVYEQMEDIIISLWQSLKPCVK 120

QY 122 MTFMCIONCTDIKN--NNTSGTENRTSSE-----NPMKTCFNTITVLKDKKEKQAL 174  
 DB 121 LTPLCVTLNCDLNTNTNNTANNNSSEGTIKGKNGKNSFNITTSIRDKQKEVAL 180

QY 175 FYVSDLTKLADNTNTMYTLINCNTTIKQCPKVPSEPIIYCAPAGYAIKCSNAE 234  
 DB 181 LYKLDIVSI---NNDSTYSLSCNTSVITQCPISPEPIHYCAPAGFAILKDKK 237

QY 235 FNGTCKSNISVVTCHGKIPVSTQILNGTLSEKTRIMGNKISDGNIIIVTLSSDI 294  
 DB 238 FSGKSGCKNVSTVQCHGIRPVSTQILLNGSLAEVYIRSENFNDNAKTIIVHLSV 297

QY 295 EITCVRPNGNQTVOEMKIGP-WAYTSM--ALGTGNSRNRVAYCOYNTTEWEKALKNTAER 351  
 DB 298 QINCTEPNKR-KRIHIGPGRAFYTKNIIGT----IRQAHCNISRAKNWDLRIVSK 352

QY 352 YLELINTEGNTMTIFNRSODGSQVETHLHFNCHGEFFYCNTEMFNTFLCNGTNC-N 410  
 DB 353 LKEQFK----NKTIVFNGS--SGDPEIVMSIFNGCGGEFFYCNTEMFNTFLCNGTNC-N 410

QY 411 NQTOSINSANGMLPCKLQVYRSMRGSGLYAPPFPGNLTICISHITGMILQ---MDAPW 466  
 DB 406 NITGSNN-NITLQCKIKQIINMVEGKATYAPPIEGQIRCSNITGLLIFDQKQDT 464

QY 467 NKTENTRPTGDMKMDIWRNELEFKYKVRVPSPVAPTPIARPVIGTGTREKRAVLGM 526  
 DB 465 NDTE-IFRPGGDMRDNWRSELYKYKVIEPLGVAPTAKRVV----QREKRA-AIGA 518

QY 527 LFLGLVSAAGTGAATATLVQTHSVKIGTVQOQNDLLRAIQAOQELRLSVNGIRQLR 586  
 DB 519 LFLGFLGAAGTGAASVTLTVQARLLUSGIVQOQNDLLRAIEAQOEMQLTVWGIKQLQ 578

QY 587 ARLLALETLIQOOLLNMGCKGRCLICTSVKWNTEWNTNINQINGNTLWQEWQOQID 646  
 DB 579 ARILAVETLKQOGLGICGSKLICITVTPWNASRN-KSLDDIWNNTWQWQEREID 637

QY 647 NVSSTIYEIQAOVQOQONCKKLELDEWASLWNLWDLTKWLYIKAIYIYVAGLIGVR 706

DB 638 NYTSLIYSLEKSTQOENNEQELLELDKWSLWNNFDTNWLWYIKIFIMVGLVGLR 697  
 QY 707 IVMTVLNVRNIRQGYQPLSLQIPTRQOSEATPGRTGEGGDEGRP---RLIPSPQGF 763  
 DB 698 IVFAVLISVNRVQGYSPLSLQTRPPVPRGPRPEGIEEGEDRDTSGRLV---HGFL 754  
 QY 764 PLYLDTLRTIILWSYHLLSNLISGTTQVISHRLGLMLGKILQKIDACRICAAYIHWLQ 823  
 DB 755 AIIWDLRLSLFLSFVHLRLDLLTAARIVE-----LLGRRGWELKYWNLLQYWSQE 807

QY 824 LOKSATSLDTFAVANWNTDDIILGILQIRLGRGILNIPRRVROGERSLL 873  
 DB 808 LASSAVSLNATDIAVAEGTDRVLEVLQRAGRALLHPTIHRQGLERALL 857

RESULT 15  
 US-08-647-714-4  
 ; Sequence 4, Application US/08647714  
 ; Patent No. 5869313  
 ; GENERAL INFORMATION:  
 ; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,  
 ; APPLICANT: GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT  
 ; APPLICANT: G.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND  
 ; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORGAN & FINNEGAN  
 ; STREET: 345 PARK AVENUE  
 ; CITY: NEW YORK  
 ; STATE: NEW YORK  
 ; COUNTRY: USA  
 ; ZIP: 10154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: FLOPPY DISK, 3.5"  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/647,714  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/388,809  
 ; FILING DATE: 15-FEB-1995  
 ; APPLICATION NUMBER: US 08/022,835  
 ; FILING DATE: 25-FEB-1993  
 ; APPLICATION NUMBER: US 07/599,491  
 ; FILING DATE: 17-OCT-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: LESLIE A. SERUNIAN  
 ; REGISTRATION NUMBER: 35,353  
 ; REFERENCE/DOCKET NUMBER: 2026-4092US2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 751-6840  
 ; TELEFAX: (212) 751-6849  
 ; TELEX: 421792  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 857 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-647-714-4

Query Match 47.8%; Score 2211; DB 2; Length 857;  
 Best Local Similarity 49.1%; Pred. No. 3.2e-173;  
 Matches 437; Conservative 159; Mismatches 240; Indels 54; Gaps 22;

QY 5 MRAMG-KRNRK--LGLIYVMAIIPCLSSSQLYATYAGVPVWEDAAPVLFCSANLIT 61  
 DB 1 MRVKGIRRYQHGWGTLGLMISATEKLMVTVYGVVPVKEATITLFCASDAKAY 60



Db 1 MRVKGIRRNQVHGWGCTMLGLLMLICSATEKULWTVYVGVVWKEATTTLFCASDAKAY 60  
QY 62 STEKHNVWASQACVPTDPPEHYELLANVTDFNFIWENYVVEQMOEDIIISLWDSLAPCIQ 121  
Db 61 DTEVHNWATHACVPDPNPQVELVNYFNFNWNANNVVEQMHEDIISLWDSLAPCVK 120  
QY 122 MTFMCOMNCTDIKN--NNTSGTENRTSSSE-----NPMKTCFENITTVLKDKKKEKQAL 174  
Db 121 LTPLCVTLNCTDLRNTTNNSTANNNSSEGTIKGEMKNCSEFNITTSIRDKMKREYAL 180  
QY 175 FYVSDLTKLADNNTWMTYLINCSTTLKQCPKVSFEPIPIYCAPAGYALFKCNSAE 234  
Db 181 LYKLDVSI---NDSYSLISNTSVITQACPFSEPIPIHYCAPAGFAILKCNCKK 237  
QY 235 FNGTGKCSNISVVTCHGKPTVSTQILNGLSKIRIMGKNISDSGKNIIVTLSSDI 294  
Db 238 FSGKSGCKNVSTVQCHGIRPVYSTQLLNGSLAEVIRSENFENDNAKTIIVHLNESV 297  
QY 295 EITCVPRGNQTVQEMKIGP--MAWYSM--ALGTGSNRSRVAYCOYNTTEWEKALKNTAER 351  
Db 298 QINCTRPYNKR-KRIHIGPGRAYTTKNIIGT---IRQAHCNISRAKWNQTLRQIVSK 352  
QY 352 YLELINNTEGNTMFINRSQDSQDVEVTHLFNCHGEFFCYCNTSEMFTFLCNGTNC-N 410  
Db 353 LKEQFK---NKTIVFNQS--SGGDPFVHMSFNGCGEFTYCNTPSPLFNSTW---NGNNTWN 405  
QY 411 NTQSIANSANGMIPCKLKQVYRWMRGSGLYAPPIGNLTCISIHITGMILQ-----MDAPW 466  
Db 406 NTGGSNN-NITLQCKIKQIINMQVEGKAIYAPPIEGQIRCSSNIYGLLLTRDGGKDDT 464  
QY 467 NKTENTFRPIGDMKDIWRNELFKYKVVYKPFSAVPTPIAREVIGTGHREKRRAVGLGM 526  
Db 465 NDTE-IFRPGGDMRNRSELYKYKYVTEPLGVAPTRAKRRV---QREKRA-AIGA 518  
QY 527 LFLGVLISAAGSTGAATALTQVTHSVIKIGIVQOQDNLLRAIOAQOQLLELSWGIQRLR 586  
Db 519 LFLGFLGAAGSTGAASVTLTVQARLLSGIVQOQNNLLRAIEAQOQMLQLTVWGIKQLQ 578  
QY 587 ARLLALETLLIONOLLNLWGKRLCYTSVKWNETWRNTNINOIWNGLTWOEDQOQID 646  
Db 579 ARILAVERYLKDQOLLGKINGCSKLICTTVPNWASWSN-KSLDDIWNNTWQWQEREID 637  
QY 647 NVSTIYEETQKAOVQOQNEKKLELDEWASLWNWLDITKWLWYIKIAIIVGALIGVR 706  
Db 638 NYTSLIYSLLEKSQTQEMNEQELLEDKWASLWNWFDITNWLWYIKIFIMIVGLVGLR 697  
QY 707 IVMIVNLVRNIRQGYOPLSLQIPTQQOSEAFPGRTGEGGDEGRP---RLIPSPQGL 763  
Db 698 IVFAVLISVNRVQGYSPLSLQTRPPVPRGDRPEGIEEGGERDRDTSGRLV---HGFL 754  
QY 764 PLYTDLRTIILWSYHLLSLISGTTVISHRLGLWILGOKIIDACRICAAVIHWLQOE 823  
Db 755 AITWDLRSLEFSYHRLDLILLIARIVE-----LGRRCWEVLKYWNWLLQYWSQE 807  
QY 824 LQKSATSLIDTFVAVANWTDIILGIQRGILNIPRVRQGFERSLL 873  
Db 808 LKSSAVSLLNATDIAVAEGTDRVIEVQRAGRAILHIPTRIOGLERALL 857

Search completed: January 25, 2001, 11:00:43  
Job time: 559 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2001, 10:51:24 ; Search time 43.17 Seconds  
(without alignments)  
2370.223 Million cell updates/sec

Title: US-08-911-824-61  
Perfect score: 4622  
Sequence: 1 MIVTPRAGKRNKRLGILYI.....GRGILNPRVRQGFERSLL 873

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTRMBL\_15:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phase:  
10: sp\_plant:  
11: sp\_prodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3859	83.5	879	12 Q9WU9	Q9WU9 human immun
2	3670.5	79.4	876	12 Q9YKB2	Q9YKB2 human immun
3	3664.5	79.3	872	12 O57072	O57072 human immun
4	3664	79.3	871	12 O57073	O57073 human immun
5	3654.5	79.1	900	12 Q9QN28	Q9QN28 human immun
6	3650	79.0	871	12 O57074	O57074 human immun
7	3632	78.6	863	12 Q77377	Q77377 human immun
8	3577.5	77.4	860	12 Q93091	Q93091 human immun
9	3414	73.9	877	12 Q76576	Q76576 human immun
10	3381.5	73.2	876	12 Q79670	Q79670 human immun
11	2383	51.6	840	12 Q9QN27	Q9QN27 human immun
12	2370	51.3	863	12 O40360	O40360 human immun
13	2356.5	51.0	860	12 Q9QRW2	Q9QRW2 human immun
14	2356	51.0	862	12 Q9WLG8	Q9WLG8 human immun
15	2342.5	50.7	859	12 Q9WJ11	Q9WJ11 human immun
16	2338.5	50.6	537	12 Q9IED8	Q9IED8 human immun
17	2334.5	50.5	861	12 Q73366	Q73366 human immun
18	2334	50.5	859	12 Q76128	Q76128 human immun
19	2332.5	50.5	864	12 Q70210	Q70210 human immun

20	2330	50.4	870	12 Q9WY6	Q9WY6 human immun
21	2328.5	50.4	536	12 Q9IEE5	Q9IEE5 human immun
22	2328.5	50.4	845	12 Q9ID89	Q9ID89 human immun
23	2324.5	50.3	855	12 Q9YU3	Q9YU3 human immun
24	2324.5	50.3	861	12 Q73360	Q73360 human immun
25	2323.5	50.3	855	12 Q70016	Q70016 human immun
26	2321.5	50.2	851	12 Q73365	Q73365 human immun
27	2319.5	50.2	853	12 Q70202	Q70202 human immun
28	2317.5	50.1	856	12 Q9WLG6	Q9WLG6 human immun
29	2317	50.1	847	12 Q9WIS1	Q9WIS1 human immun
30	2316	50.1	854	12 Q97016	Q97016 human immun
31	2316	50.1	857	12 P90249	P90249 human immun
32	2315.5	50.1	868	12 Q80225	Q80225 human immun
33	2314.5	50.1	853	12 Q9WBT5	Q9WBT5 human immun
34	2314.5	50.1	857	12 Q97065	Q97065 human immun
35	2314.5	50.1	858	12 O11945	O11945 human immun
36	2314	50.1	854	12 Q72744	Q72744 human immun
37	2312.5	50.0	851	12 Q73361	Q73361 human immun
38	2312.5	50.0	854	12 Q9YV14	Q9YV14 human immun
39	2312	50.0	857	12 Q9YU2	Q9YU2 human immun
40	2310	50.0	861	12 Q9WY9	Q9WY9 human immun
41	2308	49.9	840	12 Q9WBT0	Q9WBT0 human immun
42	2308	49.9	846	12 Q9ID98	Q9ID98 human immun
43	2306.5	49.9	849	12 Q9YU11	Q9YU11 human immun
44	2305	49.9	854	12 Q9WSE4	Q9WSE4 human immun
45	2304.5	49.9	854	12 Q9WC85	Q9WC85 human immun

## ALIGNMENTS

RESULT	1
Q9WU9	
ID	Q9WU9 PRELIMINARY; PRT; 879 AA.
AC	Q9WU9;
DT	01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE	GP160 PRECURSOR.
GN	ENV.
OS	Human immunodeficiency virus type 1.
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX	NCBI_TaxID=11676;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-VI686;
RX	MEDLINE-99223950; PubMed-10207543;
RA	Janssens W., Heyndrickx L., Van der Auwera G., Nkengasong J.,
RA	Beirnaert E., Vereecken K., Coppens S., Willems B., Franssen K.,
RA	Peeters M., Ndumbe P., Delaporte E., Van Der Groen G.,
RT	"Interpatient genetic variability of HIV-1 group O.";
RL	AIDS 13:41-48(1999).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-VI686;
RX	MEDLINE-96426454; PubMed-8828748;
RA	Delaporte E., Janssens W., Peeters M., Buve A., Dibaba G.,
RA	Perret J.L., Ditsambou V., Georges Courbot M.C., Georges A.,
RA	Bourgeois A., Samb B., Henzel D., Heyndrickx L., Franssen K.,
RA	Van Der Groen G., Larouz B., Mbe J.R.;
RT	"Epidemiological and molecular characteristics of HIV infection in
RL	Gabon, 1986-1994.";
RN	AIDS 10:903-910(1996).
DR	EMBL; X96526; CAJ65373.1; -
DR	INTERPRO; IPR000287; -
DR	INTERPRO; IPR000777; -
DR	PFAM; PF00516; GP120; 1.
DR	PFAM; PF00517; GP41; 1.
KW	Signal.
FT	SIGNAL. 1 29
FT	CHAIN 30 524
FT	CHAIN 132 156
FT	CHAIN 164 200
FT	POTENTIAL.
FT	GP120.
FT	V1.
FT	V2.

```
FT CHAIN 302 335 V3.
FT CHAIN 395 427 V4.
FT CHAIN 458 475 V5.
FT CHAIN 525 879 GP41.
SQ SEQUENCE 879 AA; 99099 MW; BC6422ADDF1A3409 CRC64;

Query Match
Best Local Similarity 82.5%; Score 3859; DB 12; Length 879;
Matches 730; Conservative 58; Mismatches 79; Indels 18; Gaps 8;

QY 1 MIVTMAMGRKRLGILYIVMALIIPCLSSSOLYATYAGVPVWEDAAPVLFCSADANL 60
DB 1 MIVTMAMGRKRLGILYIVMALIIPCLSSSOLYATYAGVPVWEDAAPVLFCSADANL 60
QY 61 TSTEKHNVASQACVPTDPTPHEYLITNTDNIENWENYVWEOQEDIIISLWDSLKPCI 120
DB 61 TSTEKHNVASQACVPTDPTPHEYLITNTDNIENWENYVWEOQEDIIISLWDSLKPCI 120
QY 121 QMTFMCIONMCTDIKNNTSGTENRTS-----SSENPMKTCFEFNITTVLKDKEKKA 173
DB 121 QMTFMCIONMCTDIKNNTSGTENRTS-----SSENPMKTCFEFNITTVLKDKEKKA 173
QY 174 LFVYSDLVK--TDN---STMVTLINCNTTIKQACPKVSFEPIPIHYCAPAGYAFKCNNT 236
DB 174 LFVYSDLVK--TDN---STMVTLINCNTTIKQACPKVSFEPIPIHYCAPAGYAFKCNNT 236
QY 234 EFNCTGKCSNISVVTCTHGIKIPVSTQLLNGTLSEKIRKINGKNISDSGKNIIVTLSSD 293
DB 234 EFNCTGKCSNISVVTCTHGIKIPVSTQLLNGTLSEKIRKINGKNISDSGKNIIVTLSSD 293
QY 237 EFNCTGKCNITVTCTHGIKIPVSTQLLNGTLSEKIRKINGKNISDSGKNIIVTLSS 296
DB 237 EFNCTGKCNITVTCTHGIKIPVSTQLLNGTLSEKIRKINGKNISDSGKNIIVTLSS 296
QY 294 IETCVRPGNNTVOEMKIGPMAYSMALGTGSRKVRAYCOYNTTEWELKALNTAERYL 353
DB 294 IETCVRPGNNTVOEMKIGPMAYSMALGTGSRKVRAYCOYNTTEWELKALNTAERYL 353
QY 297 ISMTCEPGRNH--TVQEMKIGPMAYSMALGTGSRKVRAYCOYNTTEWELKALNTAERYL 355
DB 297 ISMTCEPGRNH--TVQEMKIGPMAYSMALGTGSRKVRAYCOYNTTEWELKALNTAERYL 355
QY 354 ELINTEGNT--TMIFNRSDGSDVEYTHLHFNCHGEFFYCNTESEMPNYTFLCNGTNCNT 412
DB 354 ELINTEGNT--TMIFNRSDGSDVEYTHLHFNCHGEFFYCNTESEMPNYTFLCNGTNCNT 412
QY 356 ELVNTNSKNTVTMIFNOSNDG--DPEVTHLHFNCHGEFFYCNTESEMPNYTFLCNGTNCNT 414
DB 356 ELVNTNSKNTVTMIFNOSNDG--DPEVTHLHFNCHGEFFYCNTESEMPNYTFLCNGTNCNT 414
QY 413 QSINSANG--MIPCKLKQVVRSMRGGSLYAPPIGNLTCISHTIGMILQMDAPNKTE 470
DB 413 QSINSANG--MIPCKLKQVVRSMRGGSLYAPPIGNLTCISHTIGMILQMDAPNKTE 470
QY 415 QNNNA--HNDTQIPCRLLKQVVRSMRGGSLYAPPIGNLTCISHTIGMILQMDAPNKTE 474
DB 415 QNNNA--HNDTQIPCRLLKQVVRSMRGGSLYAPPIGNLTCISHTIGMILQMDAPNKTE 474
QY 471 NTFPRIGDMKDILWNELEFKYKVRVPKPSVAPTARVIGTGRKVRAYGLGMLFG 530
DB 471 NTFPRIGDMKDILWNELEFKYKVRVPKPSVAPTARVIGTGRKVRAYGLGMLFG 530
QY 475 TTFPRVGGDMKDILWNELEFKYKVRVPKPSVAPTARVIGTGRKVRAYGLGMLFG 534
DB 475 TTFPRVGGDMKDILWNELEFKYKVRVPKPSVAPTARVIGTGRKVRAYGLGMLFG 534
QY 531 VLSAAGSTGAATAALTVOQTHSVIKGIVQOQDNLRAIOAQOQLRLSVGIRQLRALL 590
DB 531 VLSAAGSTGAATAALTVOQTHSVIKGIVQOQDNLRAIOAQOQLRLSVGIRQLRALL 590
QY 535 VLSAAGSTGAATAALTVOQTHSVIKGIVQOQDNLRAIOAQOQLRLSVGIRQLRALL 594
DB 535 VLSAAGSTGAATAALTVOQTHSVIKGIVQOQDNLRAIOAQOQLRLSVGIRQLRALL 594
QY 591 ALETLLQOOLLNLWCKGRLCYTSVKNWETWNTNIN--QINGNLWQWDOQIDNV 648
DB 591 ALETLLQOOLLNLWCKGRLCYTSVKNWETWNTNIN--QINGNLWQWDOQIDNV 648
QY 595 ALETLLQOOLLNLWCKGRLCYTSVKNWETWNTNIN--QINGNLWQWDOQIDNV 654
DB 595 ALETLLQOOLLNLWCKGRLCYTSVKNWETWNTNIN--QINGNLWQWDOQIDNV 654
QY 649 SSTIYEIOKAQVOQOQONKLEDEWASLWNLDTIKWLYKTAIIVGALIGRIV 708
DB 649 SSTIYEIOKAQVOQOQONKLEDEWASLWNLDTIKWLYKTAIIVGALIGRIV 708
QY 655 SSTIYEIOKAQVOQOQONKLEDEWASLWNLDTIKWLYKTAIIVGALIGRIV 714
DB 655 SSTIYEIOKAQVOQOQONKLEDEWASLWNLDTIKWLYKTAIIVGALIGRIV 714
QY 709 MIVNLVRNIROGYOPLSLQIPTRQOSEATPGTGEGBRPRIPSPGFPPLLYT 768
DB 709 MIVNLVRNIROGYOPLSLQIPTRQOSEATPGTGEGBRPRIPSPGFPPLLYT 768
QY 715 MIVNLVRNIROGYOPLSLQIPTRQOSEATPGTGEGBRPRIPSPGFPPLLYT 774
DB 715 MIVNLVRNIROGYOPLSLQIPTRQOSEATPGTGEGBRPRIPSPGFPPLLYT 774
QY 769 DLRTIILWVSHLLSLISGTOVISHRLGLTILGOKIIDACRICAAVHYHLOEKSA 828
DB 769 DLRTIILWVSHLLSLISGTOVISHRLGLTILGOKIIDACRICAAVHYHLOEKSA 828
QY 775 DLRTIILWVSHLLSLISGTOVISHRLGLTILGOKIIDACRICAAVHYHLOEKSA 834
DB 775 DLRTIILWVSHLLSLISGTOVISHRLGLTILGOKIIDACRICAAVHYHLOEKSA 834
QY 829 TSLIDTFAVANWTDIIILGIORLARGILNIPRRVQGFERSLL 873
DB 829 TSLIDTFAVANWTDIIILGIORLARGILNIPRRVQGFERSLL 873
QY 835 TSLIDTFAVANWTDIIILGIORLARGILNIPRRVQGFERSLL 879
DB 835 TSLIDTFAVANWTDIIILGIORLARGILNIPRRVQGFERSLL 879
```

RESULT 2

```
O9YKB2
ID O9YKB2 PRELIMINARY; PRT; 876 AA.
AC O9YKB2;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE GP160 PRECURSOR;
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CA9;
RX MEDLINE=99223950; PubMed=10207543;
RA Janssens W., Heyndrickx L., Van der Auwera G., Nkengasong J.,
RA Belnart E., Vereecken K., Coppens S., Willems B., Franssen K.,
RA Peeters M., Ndumbe P., Delaporte E., Van Der Groen G.;
RT "Interpatient genetic variability of HIV-1 group O.";
RL AIDS 13:41-48(1999).
DR EMBL: X96522; CAA65369.1; -
DR INTERPRO: IPR000328; -
DR INTERPRO: IPR000777; -
DR PFAM: PF00516; GP120; 1.
DR PFAM: PF00517; GP41; 1.
KW Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 524 GP120.
FT CHAIN 132 156 V1.
FT CHAIN 158 199 V2.
FT CHAIN 301 335 V3.
FT CHAIN 394 425 V4.
FT CHAIN 458 474 V5.
FT CHAIN 525 876 GP41.
SQ SEQUENCE 876 AA; 98879 MW; B8E4D1B2EC656C9F CRC64;
```

Query Match 79.4%; Score 3670.5; DB 12; Length 876;  
Best Local Similarity 80.1%; Pred. No. 1.5e-295;  
Matches 707; Conservative 63; Mismatches 96; Indels 17; Gaps 12;

```
QY 1 MIVTMAMGRKRLGILYIVMALIIPCLSSSOLYATYAGVPVWEDAAPVLFCSADANL 60
DB 1 MIVTMAMGRKRLGILYIVMALIIPCLSSSOLYATYAGVPVWEDAAPVLFCSADANL 60
QY 61 TSTEKHNVASQACVPTDPTPHEYLITNTDNIENWENYVWEOQEDIIISLWDSLKPCI 120
DB 61 TSTEKHNVASQACVPTDPTPHEYLITNTDNIENWENYVWEOQEDIIISLWDSLKPCI 120
QY 121 QMTFMCIONMCTDIKNNTSGTENRTSSENPMKTCFEFNITTVLKDKEKKAQALFYVSD 179
DB 121 QMTFMCIONMCTDIKNNTSGTENRTSSENPMKTCFEFNITTVLKDKEKKAQALFYVSD 179
QY 180 LTKLADNNTN---TMYTLINCNTTIKQACPKVSFEPIPIYCAPAGYAFKCSAFBN 236
DB 180 LTKLADNNTN---TMYTLINCNTTIKQACPKVSFEPIPIYCAPAGYAFKCSAFBN 236
QY 181 LVKL--NNTGNDTMYTLINCNTTIKQACPKVSFEPIPIYCAPAGYAFKCSAFBN 238
DB 181 LVKL--NNTGNDTMYTLINCNTTIKQACPKVSFEPIPIYCAPAGYAFKCSAFBN 238
QY 237 GTGKCSNVVTCCHGKIPVSTQILNGTLSEKIRKINGKNISDSGKNIIVTLSSDIEI 296
DB 237 GTGKCSNVVTCCHGKIPVSTQILNGTLSEKIRKINGKNISDSGKNIIVTLSSDIEI 296
QY 239 GTGKCHNITVTCTHGIKIPVSTHILNGTLSEKIRKINGKNISAKNIIVTLNVTYDI 298
DB 239 GTGKCHNITVTCTHGIKIPVSTHILNGTLSEKIRKINGKNISAKNIIVTLNVTYDI 298
QY 297 TCVRPGNNTVOEMKIGPMAYSMALGTGS--NRGRVAYCOYNTTEWELKALNTAERYL 355
DB 297 TCVRPGNNTVOEMKIGPMAYSMALGTGS--NRGRVAYCOYNTTEWELKALNTAERYL 355
QY 299 TCERPGNH--TVQIRIGPLANSYMGIEKNSKNSRLAYCNVTDKRALQTAERYLEL 357
DB 299 TCERPGNH--TVQIRIGPLANSYMGIEKNSKNSRLAYCNVTDKRALQTAERYLEL 357
QY 356 INNTGNTMTIFNRSDGSDVEYTHLHFNCHGEFFYCNTESEMPNYTFLCNGTNCNTOSI 415
DB 356 INNTGNTMTIFNRSDGSDVEYTHLHFNCHGEFFYCNTESEMPNYTFLCNGTNCNTOSI 415
QY 358 INYT--NSVMTFNHISAGDDDETTVMHFNCHGEFFYCNTESEMPNYTFLCNGTNCNTOSI 415
DB 358 INYT--NSVMTFNHISAGDDDETTVMHFNCHGEFFYCNTESEMPNYTFLCNGTNCNTOSI 415
QY 416 NSANG--MIPCKLKQVVRSMRGGSLYAPPIGNLTCISHTIGMILQMDAPNKTE 472
DB 416 NSANG--MIPCKLKQVVRSMRGGSLYAPPIGNLTCISHTIGMILQMDAPNKTE 472
QY 416 TNDNGTDWIPCRLLKQVVRSMRGGSLYAPPIGNLTCISHTIGMILQMDAPNKTE 475
DB 416 TNDNGTDWIPCRLLKQVVRSMRGGSLYAPPIGNLTCISHTIGMILQMDAPNKTE 475
```

```

QY 473 FRPIGDMKDIWRNELFKYKVVVPFSPVAPTPARPVIGTGT-HREKRAVGLMFLGV 531
DB 476 LRPAGDMEDIRWTELEFKYKVVVPFSPVAPTPARPVIGTGT-PHREKRAVGLMFLGV 535
QY 532 LSAAGSTGAAATATVQTHSVKIGIVQOQDNLRAIQAOQELLRLSVGIRQLRALLA 591
DB 536 LSAAGSTGAAATATVQTHSVKIGIVQOQDNLRAIQAOQELLRLSVGIRQLRALLA 595
QY 592 LETLTONQOLLNMGCKGRLCYTSVKWNETW-RNTNTNINQWGLNLTWOEWQDQDNYSS 650
DB 596 LETLTONQOLLNMGCKGRLCYTSVKWNETW-RNTNTNINQWGLNLTWOEWQDQDNYSS 653
QY 651 TIYEIEIQAOVQOQNEKLELDEWASLWNLDTTKWLWIKIAIIVGALIGVRYMI 710
DB 654 TIYEIEIQAOVQOQNEKLELDEWASLWNLDTTKWLWIKIAIIVGALIGVRYMI 713
QY 711 VLVNVRNROGYOPLSLQIPTROQSEAEPTGRTGGGDEGRPLIPSPQGFLLPYTDL 770
DB 714 VLVNVRNROGYOPLSLQIPTROQSEAEPTGRTGGGDEGRPLIPSPQGFLLPYTDL 773
QY 771 RTIILWSYHLLSNLISGTVQTHSVKIGIVQOQDNLRAIQAOQELLRLSVGIRQLR 830
DB 774 RTIILWSYHLLSNLISGTVQTHSVKIGIVQOQDNLRAIQAOQELLRLSVGIRQLR 833
QY 831 LIDTFAVANWTDIIILGIRGLRGILNIPRRVROGFERSLL 873
DB 834 LIDTFAVANWTDIIILGIRGLRGILNIPRRVROGFERSLL 876

RESULT 3
O57072 PRELIMINARY; PRT; 872 AA.
AC O57072;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DE 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SAMPLE 276HA;
RA Hunt J.C., Golden A.M., Lund J.K., Grutler L.G., Zekeng L., Obiang J.,
RA Kaptue L., Hampel H., Vallari A., Devare S.G.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBSJ databases.
DR EMBL; U82991; AAB94314.1; -
DR INTERPRO; IPR000328; -
DR PFAM; PF00516; GP120; 1.
DR AFAM; PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 872 AA; 97767 MW; 13B84456B62B5302 CRC64;

Query Match 79.3%; Score 3664.5; DB 12; Length 872;
Best Local Similarity 77.8%; Pred. No. 4.6e-295;
Matches 690; Conservative 68; Mismatches 100; Indels 29; Gaps 8;

QY 1 MIVTRAMKRNKRKILGILYVMAIIPCL-----SSSQLYATYAGVPVWEDAAVPLFCAS 56
DB 1 MIVTRAMKRNKRKILGILYVMAIIPCL-----SSSQLYATYAGVPVWEDAAVPLFCAS 60
QY 57 DANLSTKHNVWASQAVPTDPTHEYLNTVNTFNWENWYEQMEDIIISLWDSQL 116
DB 61 DANLSTKHNVWASQAVPTDPTHEYLNTVNTFNWENWYEQMEDIIISLWDSQL 120
QY 117 KPCIQWTFMCIQMCTDIKNNNTSGTNRSSSENPMKTCENITVLKDKKQKALFY 176
DB 121 KPCIQWTFMCIQMCTDIKNNNTSGTNRSSSENPMKTCENITVLKDKKQKALFY 172
QY 177 VSDLTKLADNNTTNTMYTLINCNTTIQACPKVSEFPIPIYCAPAGVAFKCNSEFN 236

```

```

DB 173 LSDLMLKNDNTNTMYTLINCNTTIQACPKVSEFPIPIYCAPAGVAFKCNTEFN 232
QY 237 GTGCSNIVVTCFHGKIPVSTQILNGLTSLKEKIRIMKNSDSGNKIIVTSLSDIEI 296
DB 233 GTGCKNIIVVTCFHGKIPVSTQILNGLTSLKEKIRIMKNSDSGNKIIVTSLSDIEI 292
QY 297 TCVFPGNQTVQEMKIGPMWYSMALGTGSGNRSRVAVQYNTTWEKALKNTAERYLELI 356
DB 293 TCRPA--MDVQEMRIGPMWYSMALGTGSGNRSRVAVQYNTTWEKALKNTAERYLELI 350
QY 357 NTEGNTMIFNRSDGSDEYVTHLHFNCHGEFFYCNTESEMYTFLCNGNC-----NN 411
DB 351 NNT-GNISMIEN-GTSGGDPVTHLHFNCHGEFFYCNTESEMYTFLCNGNC-----NN 408
QY 412 TQNSINANGMTPCKLKQVVRSMGSGSLYAPPIGNLTCTSHLTGMILQMDAPNKTEN 471
DB 409 SSQTNSSNTFPCCLKQVVRSMGSGSLYAPPIGNLTCTSHLTGMILQMDAPNKTEN 468
QY 472 ----TFRPIGDMKDIWRNELFKYKVVVPFSPVAPTPARPVIGTGTTHREKRAVGLM 526
DB 469 GNRNATLRPVGEMKDIWRTELLNKKVRIKPFVAPTPARPVIGTGTTHREKRAVGLM 528
QY 527 LFLGVLSAAGSTGAAATATVQTHSVKIGIVQOQDNLRAIQAOQELLRLSVGIRQLR 586
DB 529 LFLGVLSAAGSTGAAATATVQTHSVKIGIVQOQDNLRAIQAOQELLRLSVGIRQLR 588
QY 587 ARLLALETLIQOQLNMGCKGRLCYTSVKWNETWNTNTNINQWGLNLTWOEWQDQID 646
DB 589 ARLLALETLIQOQLNMGCKGRLCYTSVKWNETWNTNTNINQWGLNLTWOEWQDQID 645
QY 647 NVSTIYEIEIQAOVQOQNEKLELDEWASLWNLDTTKWLWIKIAIIVGALIGVR 706
DB 646 NISSIYDEIEIQAOVQOQNEKLELDEWASLWNLDTTKWLWIKIAIIVGALIGVR 705
QY 707 IVMIVLNVNRNROGYOPLSLQIPTROQSEAEPTGRTGGGDEGRPLIPSPQGFLL 766
DB 706 IVMIVLNVNRNROGYOPLSLQIPTROQSEAEPTGRTGGGDEGRPLIPSPQGFLL 765
QY 767 YTDLRTILWSYHLLSNLISGTVQTHSVKIGIVQOQDNLRAIQAOQELLRLSVGIRQLR 826
DB 766 YTDLRTILWSYHLLSNLISGTVQTHSVKIGIVQOQDNLRAIQAOQELLRLSVGIRQLR 825
QY 827 SATSLIDTFAVANWTDIIILGIRGLRGILNIPRRVROGFERSLL 873
DB 826 SATSLIDTFAVANWTDIIILGIRGLRGILNIPRRVROGFERSLL 872

RESULT 4
O57073 PRELIMINARY; PRT; 871 AA.
AC O57073;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DE 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SAMPLE 341HA;
RA Hunt J.C., Golden A.M., Lund J.K., Grutler L.G., Zekeng L., Obiang J.,
RA Kaptue L., Hampel H., Vallari A., Devare S.G.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBSJ databases.
DR EMBL; U82992; AAB94315.1; -
DR INTERPRO; IPR000328; -
DR PFAM; PF00516; GP120; 1.
DR AFAM; PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 871 AA; 97968 MW; 38E4DFA719E70C17 CRC64;

```







Db 1 MIVTKAMERKRNKMLTLYLALMALITPCLSLQIATYATYAGVPVWEDATPVVLCASDANL 60  
QY 61 TSTEKNHNVASQACVPTDPTPHBYLLTNTVDFNENWENYVWQEQEDIIISLWQSLKPCI 120  
Db 61 TSTEKNHNVASQACVPTDPTPHBYLLTNTVDFNENWENYVWQEQEDIIISLWQSLKPCI 120  
QY 121 QMTFMCIONCTDIKNNTSGTENRTSSSENPMKTCFENITTVLKDKEKKAALFYVDL 180  
Db 121 QMTFMCIONCTDIKNNTSGTENRTSSSENPMKTCFENITTVLKDKEKKAALFYVDL 180  
QY 181 TKLADNNTN---TWYTLNCSNITIKQACPKVSEPIPIYCAPAGYAFKCNSEFNGTG 236  
Db 171 MELNETSSNKNKNSMTNITNCSNITIKQACPKVSEPIPIYCAPAGYAFKCNSEFNGTG 236  
QY 237 GTGKCSNISVCTTGKIPVSTQILNGLTSLKSKIRMAKDIILEGKNIIVTLNGLN 290  
Db 231 GTGCRNITVCTTGKIPVSTQILNGLTSLKSKIRMAKDIILEGKNIIVTLNGLN 290  
QY 297 TCVRPNNOTQEMKIGPMWYSMAL-GTGSNRVAYCOVNTFWEKALKNTAERYL 355  
Db 291 TCERP-QIDQEMKIGPMWYSMGIGTAGNSSRAAYCKYATDMCKILKQTAERYL 348  
QY 356 INNTGNTTMIKNSQSDGVETHLHFNCHGEFFYCNTESEMYTFLCNGTNCNTQSI 415  
Db 349 VNNT-GSINMTFHS-SGDLVTHLHFNCHGEFFYCNTESEMYTFLCNGTNCNTQSI 415  
QY 416 NSANGMIPCKLKQVYRSMWGGGLYAPPVGNLTCTISHITGMLQMDAPWNTEN--TF 473  
Db 407 QGNNGTLCPKQURVYRWIRGSLYAPPVGNLTCTISHITGMLQMDAPWNTEN--TF 473  
QY 474 RPIGDMKDIDWNLKFKYVYKVPFVAPTRPVIGTTHREKRAVGLMFLVGLVS 533  
Db 467 RPIGDMKDIDWNLKFKYVYKVPFVAPTRPVIGTTHREKRAVGLMFLVGLVS 526  
QY 534 AAGTWGAATALTVPVSHVIGVQOQDNLRAIAQOQELLRLSVGIRQLRALLALE 593  
Db 527 AAGTWGAATALTVPVSHVIGVQOQDNLRAIAQOQELLRLSVGIRQLRALLALE 593  
QY 594 TLQOQQLLWCKGRLCYTSVKWNETWRNTNTNQTNWGLTQWEDQOQIDNYSSTIY 653  
Db 587 TLQOQQLLWCKGRLCYTSVKWNETWRNTNTNQTNWGLTQWEDQOQIDNYSSTIY 643  
QY 654 EETQKAOQOQEQNEKKLLEDEWASLWNLWDITKWLWIKIATVIGALIGVIMVILN 713  
Db 644 EETQKAOQOQEQNEKKLLEDEWASLWNLWDITKWLWIKIATVIGALIGVIMVILN 703  
QY 714 LVNIRGQOPLSLQIPTRQOQSEATPGTGEGGDEGRPLIPSPQGFLLPLTDLRTI 773  
Db 704 LVNIRGQOPLSLQIPTRQOQSEATPGTGEGGDEGRPLIPSPQGFLLPLTDLRTI 763  
QY 774 ILWYHLLSNLSGRTQVISHRLGLWILGOKIIDACRICAAVHYWLOELQKSAATSLID 833  
Db 764 ILWYHLLSNLSGRTQVISHRLGLWILGOKIIDACRICAAVHYWLOELQKSAATSLID 823  
QY 834 TFVAVANWTDIIIGIORLGLGILNIPRRVQFERSLL 873  
Db 824 TLAVAVANWTDIIIGIORLGLGILNIPRRVQFERSLL 863  
RESULT 8  
ID O93091 PRELIMINARY; PRT; 860 AA.  
AC O93091.  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE ENVELOPE GLYCOPROTEIN.  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-SAMPLE 193HA;  
RA Hunt J.C., Golden A.M., Lund J.K., Gurtler L.G., Zekeng L., Obiang J.,  
RA Kapteue L., Hampl H., Vallari A., Devare S.G.;  
RT "Envelope Sequence Variability and Serologic Characterization of HIV-1  
RT Group O Isolates from Equatorial Guinea."  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U82990; AAB94313.1;  
DR INTERPRO; IPR000328;  
DR INTERPRO; IPR000777;  
DR PFAM; PF00516; GP120; 1.  
DR PFAM; PF00517; GP41; 1.  
KW Envelope protein.  
SQ SEQUENCE 860 AA; 96914 MW; A24B989AAE46DD8D CRC64;

Query Match 77.48; Score 3577.5; DB 12; Length 860;  
Best Local Similarity 76.68; Pred. No. 7.3e-288;  
Matches 671; Conservative 88; Mismatches 98; Indels 19; Gaps 7;

QY 1 MIVTKAMERKRNKMLTLYLALMALITPCLSSQIATYATYAGVPVWEDATPVVLCASDANL 60  
Db 1 MIVTKAMERKRNKMLTLYLALMALITPCLSSQIATYATYAGVPVWEDATPVVLCASDANL 60  
QY 61 TSTEKNHNVASQACVPTDPTPHBYLLTNTVDFNENWENYVWQEQEDIIISLWQSLKPCI 120  
Db 61 TSTEKNHNVASQACVPTDPTPHBYLLTNTVDFNENWENYVWQEQEDIIISLWQSLKPCI 120  
QY 121 QMTFMCIONCTDIKNNTSGTENRTSSSENPMKTCFENITTVLKDKEKKAALFYVDL 180  
Db 121 QMTFMCIONCTDIKNNTSGTENRTSSSENPMKTCFENITTVLKDKEKKAALFYVDL 180  
QY 181 TKLADNNTN---TWYTLNCSNITIKQACPKVSEPIPIYCAPAGYAFKCNSEFNGTG 229  
Db 174 IKLEN-----STWYTLNCSNITIKQACPKVSEPIPIYCAPAGYAFKCNSEFNGTG 229  
QY 241 CSNISVVTCTGKIPVSTQILNGLTSLKSKIRMAKDIILEGKNIIVTLNGLN 289  
Db 230 CSNISVVTCTGKIPVSTQILNGLTSLKSKIRMAKDIILEGKNIIVTLNGLN 289  
QY 301 PGNNQVQEMKIGPMWYSMALGTGNSRVAYCOYNTTEWEKALKNTAERYL 360  
Db 290 PAIN--VQOQGLPLAVSYSLGVENITSRTAYCEYNTAAWERTLOQTAERWELVNTK 347  
QY 361 GNTMTMFNSODGSVEVTHLHFNCHGEFFYCNTESEMYTFLCNGTNCN-NTOSINSAN 419  
Db 348 INVSIIFNOSGGGDAETHLHFNCHGEFFYCNTESEMYTFLCNGTNCN-NTOSINSAN 407  
QY 420 GMIPCKLKQVYRSMWGGGLYAPPVGNLTCTISHITGMLQMDAPWNTEN--TF 477  
Db 408 GMIPCKLKQVYRSMWGGGLYAPPVGNLTCTISHITGMLQMDAPWNTEN--TF 477  
QY 478 GDMKIDWNLKFKYVYKVPFVAPTRPVIGTTHREKRAVGLMFLVGLVS 537  
Db 468 GDMKIDWNLKFKYVYKVPFVAPTRPVIGTTHREKRAVGLMFLVGLVS 527  
QY 538 TNGAATALTVPVSHVIGVQOQDNLRAIAQOQELLRLSVGIRQLRALLALE 597  
Db 528 TNGAATALTVPVSHVIGVQOQDNLRAIAQOQELLRLSVGIRQLRALLALE 587  
QY 598 NOQLNLWCKGRLCYTSVKWNETWRNTNTNQTNWGLTQWEDQOQIDNYSSTIY 657  
Db 588 NOQLNLWCKGRLCYTSVKWNETWRNTNTNQTNWGLTQWEDQOQIDNYSSTIY 644  
QY 658 KAOVQOQEQNEKKLLEDEWASLWNLWDITKWLWIKIATVIGALIGVIMVILN 717  
Db 645 KAOVQOQEQNEKKLLEDEWASLWNLWDITKWLWIKIATVIGALIGVIMVILN 704  
QY 718 IROGYQPLSLQIPTRQOQSEATPGTGEGGDEGRPLIPSPQGFLLPLTDLRTI 777  
Db 705 IROGYQPLSLQIPTRQOQSEATPGTGEGGDEGRPLIPSPQGFLLPLTDLRTI 764  
QY 778 YHLSNLISGRTQVISHRLGLWILGOKIIDACRICAAVHYWLOELQKSAATSLID 837  
Db 778 YHLSNLISGRTQVISHRLGLWILGOKIIDACRICAAVHYWLOELQKSAATSLID 837



Db	121	EkKtFLcVomNCvDLQnKT-GllNFTI---NEMrNCSENVVTLTDKKEQKQALFYSD	176
Qy	180	LTKLADNNNTN-TWYTLINCNSTIIKQACPkvSFEPiPIYyCAPAGyAfKNCsAEfNGT	238
Db	177	LSKvDNNNAVGTYYMLTNCNSTIIKQACPkvSFEPiPIHycAPGTyAfKNCDFDfNGT	236
Qy	239	GKCSNISVVTCTHGikPTvSTQILNGTLsKekIRiMGKNIsgDKsNIIVTLSSDIEITC	298
Db	237	GLCHNISVVTCTHGikPTvSTQILNGTLsREKIRiMGKNITsAKNIIVTLNTPINNCT	296
Qy	299	VRGNNQTVQEMKIGKPMANYSMALGTGSN---RSRVAYCOYNTTWEKALKNTARYLEL	355
Db	297	IREGIAE-VODIYTGPMRWRSMTLKRSNNTSPRSVAYCTYKNTVWENALQQTALRIYNL	355
Qy	356	INTEGNTMIFNRSGDGSVEYTHLHFNCHEGFEFFCYNTSEMFNTFL-CNGTNCNNTQS	414
Db	356	VNOTE-NVTIIFSR-TSGDAEYVHLHFNCHEGFEFFCYNTSGMENYTFINCTKSGCOEIKG	413
Qy	415	IN--SANGMITPCKLKQVVRSMRGSGGLYAPPTPGMLTCIshITGMILQMDAPWNT-EN	471
Db	414	SNETNKGITPCKLRQLVRSMWKGESRIYAPPTPGMLTCHSITGMILQDQPMWNTGEN	473
Qy	472	TFPIGDKMDIWRNELFKYKVVVRPFESVAPTPiARPVIGT-HREKRAYGLGMLFLG	530
Db	474	TLRPVGDKMDIWRTELKYNKVVQIAPFSVAPTKMSRPINHTPREKRAYGLGMLFLG	533
Qy	531	VLSAAGTGMGAATALTVOThSVIKGIvQOQDNLRAIOAQOELRLSVGIQLRLARLL	590
Db	534	VLSAAGTGMGAATALTvrThSVLKGIVQOQDNLRAIOAQOHLRLSVGIQLRLARLQ	593
Qy	591	ALETLQNOQLLWNGCKGRLCYTSVKKNETWRNTINQINGLNTWQEWDOQIDNYS	650
Db	594	ALETLQNOQLLWNGCKGLCYTSVKWNTSWSGRYNDSDIWNLTWQWQDQHINNYS	653
Qy	651	TIYEEIQKAQVQOEKLELDEWASLWNMLDITKWLWYIKAIiIVGALIGIRVIM	710
Db	654	IYDEIQAAQDQEKVKALLELDEWASLWNWFDITKWLWYIKAIiIVGALIGIRVIM	713
Qy	711	VLNLVRNIROGYOPLSLQIPTROSEAPGRTGEGGDEGRPLTPSPQGLPLLYTDL	770
pb	714	ILNLVRNIROGYOPLSLQIIPVHQSAETPGRTGEGGDEGRKWTALPPGLQYLITL	773
Qy	771	RTILWSYHLLSNLSIGTQTVISHLRGLWILGQKIIDACRICAAViHYWLQELQKSATS	830
Db	774	RTIILWYHLLSNLSIGRLLDYLGLGLWILGQKTEACRLCGAVMQWLQELKNSATN	833
Qy	831	LIDTFAVANWTDIILGTLQRLGRGILNTPRRVROGFSRL	873
Db	834	LLDTIAVSANWTDGIILGLQIRGQFLHPRIIRROGAERILV	876
RESULT 11			
Q9QN7	Q9QN7	PRELIMINARY;	PRT; 840 AA.
AC	Q9QN7;		
DT	01-MAY-2000 (TEMBLrel. 13, Created)		
DT	01-MAY-2000 (TEMBLrel. 13, Last sequence update)		
DT	01-JUN-2000 (TEMBLrel. 14, Last annotation update)		
DE	ENV PROTEIN.		
EN	ENV.		
OS	Human immunodeficiency virus type 1.		
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.		
OX	NCBI_TaxID=11676;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-95ML84;		
RA	Montavon C., Bibollet-Ruche F., Robertson D., Koumare B., Mulanga C.,		
RA	Esu-Williams E., Youre C., Mboup S., Saman E., Delaporte E.,		
RA	Peters M.;		
RT	The identification of a complex A/G/I/J recombinant HIV-1 virus in		
RL	different west African countries.;		
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.		

RESULT 12  
O40360 PRELIMINARY; PRT; 863 AA.  
AC O40360;  
DT 01-JAN-1998 (TREMELrel. 05, Created)  
DT 01-JAN-1998 (TREMELrel. 05, Last sequence update)  
DT 01-MAY-2000 (TREMELrel. 13, Last annotation update)  
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=FCRNP;  
RX MEDLINE=98216834; PubMed=9557756;  
RA Blanchard A., Ferris S., Chamarret S., Guetard D., Montagnier L.;  
RT "Molecular evidence for nosocomial transmission of human  
RT immunodeficiency virus from a surgeon to one of his patients."  
J. Virol. 72:4537-4540(1998).  
RL EMBL; U85914; AAC59351.1; -  
DR INTERPRO: IPR000328; -  
DR INTERPRO: IPR000777; -  
DR PFAM: PF00516; GP120; 1.  
DR PFAM: PF00517; GP41; 1.  
KW Envelope protein.  
FT MON\_TER 1  
SQ SEQUENCE 863 AA; 97628 MW; 4F635052B7397BDB CRC64;

Query Match 51.3%; Score 2370; DB 12; Length 863;  
Best Local Similarity 52.3%; Pred. No. 9.5e-188;  
Matches 464; Conservative 140; Mismatches 238; Indels 46; Gaps 14;

QY 5 MRANKRNRKL---GILYVVALIIPCLSSQLYATVYAGVPVWEDAPVLCASDANLT 61  
DB 1 VRGMORNRKGLNGILY--LGILICNAADNLWTVYGVVWMDAKTTLFCASEAKAY 58  
QY 62 STEKHNVASOACVPTDPTPEHYLLTNVDFNFIWENWVQMDIISLWDQSLKPCIQ 121  
DB 59 ETEVHNWATACVPTDNPQELIYVNTFNFNWMDNMDVQMDIISLWDQSLKPCVK 118  
QY 122 MTFQICMNCNCTDIKNNTSGTENKTS-----SENPKTCFENITTVLKDKREKK 171  
DB 119 LTPCLVTLNCNITKSAFANDINSTSAPLTVIISSEARRELTCSPNVTTELROKTKV 178  
QY 172 QALFVSDLTKLADNNTN-----TMYTLINCNTTQACPKVSFEPIPIYCAPAGYA 226  
DB 179 YSIFVQLDVQDENNSNSNSYFRLINCNASTITQACPKVSFEPIPIYCAPAGYA 238  
QY 227 IFKNSAEFNGTGCNSISVVTCTHGKPTVSTOLILNGTSLSKKIRTMGNISDSGNI 286  
DB 239 ILKCKDTFNGTGPCSNVSTVCTHGKIPVSTQLLNGSLAEKIMRTNKRSDSPGNI 298  
QY 287 IVTLSSDIEITCRVPGNNTQVEMKIGP--MAWYSMALGTGNSRVRVAYCOYNTTEWEKAL 345  
DB 299 IVQLIKPKVINCTRP--NNTRKSVHLGPGQAFY--ATGDIIGNVRQAFCTVNRTEKNLT 355  
QY 346 KNTAERYLELINTNEGNTMTFNRSQDSGVVEVHLHFNCHGEFFYCNTEMFNYTLGN 405  
DB 356 FNVSQLKEYNKTE-----YFKESSGGDVEITTHSPNCRGEFFYCNTEGLFNSSFIQ 410  
QY 406 GYNCHNTOSINSANGMIPKLVQVRSWNRGSGLYAPPINGLNCTICSHITGMILQMDAP 465  
DB 411 GTSNDND---TSANTTIPCRKIQIVRMQRVQGMIAPIPGVITCVSNITGLLTGDGE 467  
QY 466 WNKTEFTPIGGMDKIDWRNELFKYKVRVYKVPFSAVPTPIARVICTGTGREKRAYGLG 525  
DB 468 KNGINETRPIGGNWRNWSRLRYKVVRIEPLGVAPTRARRVV---BREKRAYGLG 523  
QY 526 MFLGVLSAAGSTGGAATALTVOHSHVIGVQOQDNLRAIQAOQELLRLSVWGIQL 585  
DB 524 AVFLGFLGAGSTGGAASITLTVOARQLLSVQOQSNLLRAIEAQQOMLRLTVWGIQL 583

QY 586 RARLALLETLIQOQLNLWGCKGRLYCTSVKNETRTNTNINQIWNLTQWEMDQOI 645  
DB 584 QARVLAVERYLKQQLLGIGWCGSRLLCTTAVPWNSSWN--KSFNELWNTWLOWDKEI 642  
QY 646 DNVSTIYBIEIQAKVOQEQNEKKLLEDEWASLWNLDTIKWLWIKIAIIVGALIGV 705  
DB 643 SNYSETIYRLIEDSQNQEKNEQDLSLDQWGLWNFDTIRWLWYIKIFIMVGGIIGL 702  
QY 706 RIVMVLNLVRNIRQGYQPLSLQIPTRQOSEAETPGRTGEGGDEGPRPLIPSPQGLPL 765  
DB 703 RIIFAVLSIVNRVQGYSPLSFTLIPSPRGDPREIEERGEGQDGRGIRLVNGFLAL 762  
QY 766 LYTDLRTIILSYHLLSN--LISGTQTVISHRLGLWLTGKIIDACRICAIAHYWLQEL 824  
DB 763 AWEDLRNCLFSYRLLRDLALLAARTLDRLR--GSM-----EILKYLGNLQVWQEL 814  
QY 825 QKSATSIDTFVAVANWTDIILGQRLGRLGILNIPRRVQGFERSL 872  
DB 815 RNSAISLLNTAIVAEGRDRIEILQIRAGRAILHIPRIRROGAERIL 862  
RESULT 13  
Q9QRW2 PRELIMINARY; PRT; 860 AA.  
AC Q9QRW2;  
DT 01-MAY-2000 (TREMELrel. 13, Created)  
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMELrel. 14, Last annotation update)  
DE ENV.  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Laukkanen T., Salminen M.O.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF107771; AAD49789.1; -  
DR INTERPRO: IPR000328; -  
DR INTERPRO: IPR000777; -  
DR PFAM: PF00516; GP120; 1.  
DR PFAM: PF00517; GP41; 1.  
SQ SEQUENCE 860 AA; 97757 MW; 2923610F7D7ACAC9 CRC64;

Query Match 51.0%; Score 2356.5; DB 12; Length 860;  
Best Local Similarity 51.8%; Pred. No. 1.2e-186;  
Matches 461; Conservative 146; Mismatches 232; Indels 51; Gaps 16;

QY 5 MRANKRNRKLGL---YIVMALIIPCLSSQLYATVYAGVPVWEDAPVLCASDANLT 61  
DB 1 MRVMTGORNCHLLNMGIMILGIIICSTAEENLWTVYGVVWMDAETTLFCASDAKAY 60  
QY 62 STEKHNVASOACVPTDPTPEHYLLTNVDFNFIWENWVQMDIISLWDQSLKPCIQ 121  
DB 61 EKEVHNWATACVPTDNPQELIYVNTFNFNWMDNMDVQMDIISLWDQSLKPCVQ 120  
QY 122 MTFQICMNCNCTDIKNNTSGTE--NRTSSSNPMKTCFENITTVLKDKREKQALFYVSDL 180  
DB 121 LTPCLVTLNCNNTANTNSTSANLTDVSGEMRNCNFTITELRDKKRVSYLFYKLDI 180  
QY 181 TKLADN-----NTNTMTYLNCNNTTITQACPKVSFEPIPIYCAPAGYAFK 230  
DB 181 VKINKNSFRGNSSGSSDRYLINCNTSAITQACPKVSFEPIPIYCAPAGYAFK 240  
QY 231 NSAEFNGTGCNSISVVTCTHGKIPVSTOLILNGTSLSKKIRTMGNISDSGNIIVTL 290  
DB 241 NEDEFNGTGCNVSIVCTHGRPVVSTOLLNGSLAKEEVRIRSENISDNKTIIVQF 300  
QY 291 SSDIETVCRPGNNTQVEMKIGP--MAWYSMALGTGNSRVRVAYCOYNTTEWEKALNTA 349

```
Db 301 TKPVEIICTRP--NNNTRKSRIGPQAFYGMGDIIGDIRK--AHCNVSRSKWNETLAKVA 357
QY 350 ---EYLELINTGNTMIENRSDGSDVEVTHLHFNCHEFFCYNTSEMFNYTELCNG 406
Db 358 IOLRKY-----WNTIIIFTN--SGDLEITHSFNCGGEFFCYNTSGLFNSWEND 408
QY 407 TNCN-NTOSINSANGMIPCKLKQVVRSMRGSGLYAPPPIPNGLTICSHITGMILQMD-A 464
Db 409 TKVNYTESNDIT--LOCRIKOIINMQRGQATYAPPPIGVQCRSNIITGLLTRDGG 466
QY 465 PNWKTEN--TFRPIGDMKDIWRNELFKYKVVVRPFSVAPTPIARPVIGTTHREKRAVG 523
Db 467 VTNNNTNETFRPGGDMRDNWRSELYKYVKLEPLGVAPTAKARRVY-----KREKRAVG 522
QY 524 LGMLFLGVLSSAGSTMGAATAATVQTHSVIKGIVQODNLLRAIQAOQELLRLSVGIR 583
Db 523 LAAPFFGLGAGSTMGAASITITVQARQLSGIVQOONLLRAIEAQOHLRLTVWGIR 582
QY 584 QLRLALALETLIONQOOLLNLWGCKGRICYTSVKWNETWRTNTNIQINGNLTWQEWDO 643
Db 583 QLQARVMAVERLADQOLLGWCSGKIICITAVPNWSTWSN--KSYTQIWNMTWLQWDK 641
QY 644 QIDNVSSTIYEIEOKAQVQOQONKEKLELDEWASLNNWLDITKWLWIKIAIIVGALI 703
Db 642 EISNYTDIIYQIIEESQYQOEKNEKLELDEKWANLWDFDISNWLMTIKIFIMIVGGLI 701
QY 704 GVRIVMVLNVRNITRGYOPLSLOIPTROQSEAEPTGRTGEGGDRGPRPLIPSGFL 763
Db 702 GLRIVFVLTVIKRVOGYSPLSQIHTPSPDRPRGRIEEGGDRGRIKSLVSGFL 761
QY 764 PLYTDLTILWYHLLSNISIGTQTVISHRLGLWMLGKIIDACRICAAVHYHLOE 823
Db 762 ALANDLRLSCFSYHRLRDISIATRIVELLGGOLKYLGNLL-----YWIRES 810
QY 824 LOKSATSILDTFAVANWNTDDIILGQRLGRIILNIPRRVQGFERSLL 873
Db 811 LKISALSLEDTAIVAGWTDVIEIGRIGRAILHIPRIRQGFERALL 860

RESULT 14
Q9WLJ8 PRELIMINARY; PRT; 862 AA.
AC Q9WLJ8;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N11144;
RA McCutchan F.;
RT "Limited diversity of subtype E envelope sequences from recent,
RT seroincident cases of human immunodeficiency virus type 1 infection in
RT Thailand.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF070703; AAC99347.1; -
DR INTERPRO; IPR000328; -
DR INTERPRO; IPR000777; -
DR PFAM; PF00516; GP120; 1.
DR PFAM; PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 862 AA; 97667 MW; 256A6E2C8069FC15 CRC64;

Query Match 51.08; Score 2356; DB 12; Length 862;
Best Local Similarity 51.68; Pred. No. 1.4e-186;
Matches 451; Conservative 144; Mismatches 229; Indels 50; Gaps 15;

QY 20 IVMALIIPCLSSQLYATVYAGVPWEDAAPVLCASDANLTSTKHNWNASQACVPTDP 79
```

```
Db 19 LILGLVICSASNLWTVYVGVVWRDADTTLFCASDAKAKAHETEAHNIWATHACVPTDP 78
QY 80 TPHEYLITNTVDFNFWENWVQEQMEDIIISLDQSLKPCIOQMTFMCIONCNDIKNNNT 139
Db 79 NPOEIHLENTFNWKNWVQEQMEDIVISLDQSLKPCVLTCLVTLNCDSDAKFKV 138
QY 140 SGTENRTSSSEN--PMKTCENITTVLKDKEKKKQALFYVSDTLTKLADNNTNTMYTLI 196
Db 139 STINDSNIIGNITDEVNCTFNMTTELKDKKQNVHAFYKLDIVQL--ENNNSREYRLI 197
QY 197 NCNSTTIKQACPKVSEPIPIYCAPAGYAFKCSAEFNGTGKCSNISVVTCTHGIKPT 256
Db 198 NCNSTVIKQACPKISDPPIHYCTPAGYAILKCNKDNFNGTGPCNKSVSQCTHGKIPV 257
QY 257 VSTQILLNGTSLSEKTRIMGNKNSDSGKNIIVLSSDIEITCVRPNGNQVQEMKIGP-M 315
Db 258 VSTQILLNGTSLSEKTRIMGNKNSDSGKNIIVLSSDIEITCVRPNGNQVQEMKIGP-M 316
QY 316 AWYSMALGTGNSRVAICYNTTEWEKALKNTAERYLELINNTGNTTIFNRSQDGS 375
Db 317 VFRTGDIIGDIRK--AYCEVNGTKWNEILKQVTEKKEHFNT-----IVQPPSGGD 368
QY 376 VEYTHLHFNCHEFFCYNTSEMFNYTFL-----CNGTNCNNTQINSANGMIPCKLKQ 428
Db 369 LEITHMHFNCGEFCYNTSRLENFTWIGNETMERCNGTGCNGTI-----ILPCKIKQ 421
QY 429 VWSWRGSGSLYAPPIPNGLTICSHITGMILQMDAPWKTEN--TERPIGDMKDIWRNE 487
Db 422 IINMOWAGAGYAPPPIRSGISCVSNITGILLTRDGGANNASNETFRPGGNIKDNWRSE 481
QY 488 LFYKVVVRPFSVAPTPIARPVIGTTHREKRAVGLMFLGVLSSAGSTMGAATAALT 547
Db 482 LYKVVVQIEPLGVAPTAKARRVY-----QREKRAVGIGAMIFGLNAGSTMGAATILT 537
QY 548 VQTHSVIKGIVQOONLLRAIQAOQELLRLSVGIRQLRRLALETLIONQOOLLNLWGC 607
Db 538 VQARQLSGIVQOONLLRAIEAQOHLQLTVYGIKQLQARVLAVERYLKNQKFLGLWGC 597
QY 608 KGRICYTSVKWNETWRTNTNIQINGNLTWQEWDOQIDNVSSTIYEIEOKAQVQEQNE 667
Db 598 SGKIICTAVPNWSTWSN--RSFEEIWNNTWIEWEREISNYTSQIYEILTESQNOODRNE 656
QY 668 KLELDEWASLNNWLDITKWLWIKIAIIVGALIGVRIVMVLNVRNIRGYPPLSL 727
Db 657 KDLLELDKASLNNWDFITNWLWYIKIFIMIVGGLIGLRIIFAVLSIVNVRGYSPLSL 716
QY 728 QIPTROQSEAEPTGRTGEGGDRGPRPLIPSPQGFPLPYDTRTILWYHLLSN-LIS 786
Db 717 QTPTHQREPDPRPERIEEGGEGGQDRSVRLVSGFLSLAWDDIRSLCLFLYHRLDFILI 776
QY 787 GTQTV-----ISHRLGLWMLGKIIDACRICAAVHYHLOEQLQKSLDITFAVAV 839
Db 777 ATREVELLGHSSSLKGLRRG-W-----EGLKYLKLLIYWGQELKISALSILDATAIV 828
QY 840 ANWTDIILGILQILGRIILNIPRRVQGFERSLL 873
Db 829 AGWTDVIEVQAGVWRAILHIPRIRQGFERALL 862

RESULT 15
Q9WLJ1 PRELIMINARY; PRT; 859 AA.
AC Q9WLJ1;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE ENVELOPE GLYCOPROTEIN.
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
```



